



**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CAMPUS I  
CENTRO DE CIÊNCIAS E TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA  
CURSO DE BACHARELADO EM ESTATÍSTICA**

**LUANA PINTO VIEIRA**

**MODELOS CONVENCIONAIS E DE MACHINE LEARNING NA  
PREDIÇÃO DO CRESCIMENTO DE BULBOS DE CEBOLA**

**CAMPINA GRANDE - PB  
2024**

**LUANA PINTO VIEIRA**

**MODELOS CONVENCIONAIS E DE MACHINE LEARNING NA  
PREDIÇÃO DO CRESCIMENTO DE BULBOS DE CEBOLA**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

**Área de concentração:** Estatística

**Orientador(a):** Profa. Dra. Ana Patricia Bastos Peixoto de Oliveira

**Coorientador(a):** Profa. Ma. Débora de Sousa Cordeiro

**CAMPINA GRANDE - PB**

**2024**

É expressamente proibida a comercialização deste documento, tanto em versão impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que, na reprodução, figure a identificação do autor, título, instituição e ano do trabalho.

V658m Vieira, Luana Pinto.

Modelos convencionais e de machine learning na predição do crescimento de bulbos de cebola [manuscrito] / Luana Pinto Vieira. - 2024.

39 f. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2024.

"Orientação : Prof. Dra. Ana Patricia Bastos Peixoto de Oliveira, Departamento de Estatística - CCT".

"Coorientação: Prof. Ma. Débora de Sousa Cordeiro, Departamento de Estatística - CCT".

1. Modelos não lineares. 2. Regressão logística. 3. Aprendizado de máquina. I. Título

21. ed. CDD 631.563

LUANA PINTO VIEIRA

MODELOS CONVENCIONAIS E DE MACHINE LEARNING NA  
PREDIÇÃO DO CRESCIMENTO DE BULBOS DE CEBOLA

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Estatística.

Área de concentração: Estatística

Aprovado em: 21/ 11/ 2024.

BANCA EXAMINADORA



Profa. Ma. Débora de Sousa Cordeiro (Coorientador)  
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Ricardo Alves de Olinda (Examinador)  
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Silvio Fernando Xavier Junior (Examinador)  
Universidade Estadual da Paraíba (UEPB)

“Dedico a Deus, á  
minha família, aos  
meus amigos e colegas  
de sala por todo  
incentivo.”

## AGRADECIMENTOS

Primeiramente, expresso minha profunda gratidão a Deus por ser minha fonte constante de força, determinação e inspiração em meu percurso acadêmico ao longo dos anos na universidade.

À minha família, deixo um agradecimento especial. A minha irmã Letícia e ao meu irmão Lucas, que me acompanharam de perto em cada etapa dessa jornada, agradeço por me inspirarem a ser mais confiante e determinada. À minha mãe, Marizeth, e ao meu pai, Luiz, que sempre acreditaram no meu potencial, dedico toda a minha gratidão, apoio e incentivo que foram fundamentais para que eu chegasse até aqui. A Emanuel sou eternamente grata pelo apoio constante e pela paciência.

Aos meus amigos Albert, Matheus, Ana Beatriz, Ramon, Amauri, Ingrid, Luiz e Renally, agradeço por serem uma fonte constante de motivação, por suas palavras de encorajamento, pela compreensão nos momentos difíceis e por cada risada compartilhada, que trouxe leveza aos intensos dias de estudo. Agradeço também a todos os amigos que, direta ou indiretamente, colaboraram para que este momento fosse possível, seja nos momentos de descontração, seja nos desafios compartilhados.

À minha orientadora Prof<sup>a</sup>. Dra Ana Patricia Bastos Peixoto e a minha coorientadora Prof<sup>a</sup>. Ma. Débora de Sousa Cordeiro dedico minha gratidão mais sincera pela paciência, dedicação e pelas valiosas orientações que me guiaram ao longo deste trabalho. Suas palavras de sabedoria e encorajamento foram não só fundamentais para o desenvolvimento deste projeto, mas também se tornaram inspirações pessoais que levarei comigo sempre.

Agradeço também aos membros da banca examinadora o Prof. Dr. Ricardo Alves de Olinda e o Prof. Dr. Silvio Fernando Xavier Junior pelas contribuições essenciais que enriqueceram este trabalho, oferecendo tempo, conhecimento e perspectiva para tornar este projeto mais completo e significativo.

Aos professores do Departamento de Estatística da Universidade Estadual da Paraíba que foram responsáveis por formar a base do meu conhecimento e profissionalismo, expresso minha profunda gratidão pelo apoio e confiança.

E a todos que, de alguma forma, direta ou indiretamente, contribuíram para a realização deste trabalho, fica aqui registrado o meu mais sincero e profundo agradecimento.

“O crescimento sustentável vem da melhoria contínua, que deve ser medida e analisada.”  
(Deming, 1986)

## RESUMO

Esse estudo investiga o uso de modelos de regressão não linear e algoritmos de Aprendizado de Máquina para descrever o crescimento do bulbo da cebola ao longo do tempo, visando otimizar a produção agrícola. Utilizando dados experimentais foram ajustados três modelos de crescimento não lineares: *Gompertz*, *Von Bertalanffy* e Logístico, ao peso seco do bulbo da cebola, com o tempo de crescimento como variável explicativa. A análise foi realizada no *software* R, empregando critérios de seleção como AIC (Critério de Informação de Akaike) e BIC (Critério de Informação Bayesiano) para identificar o modelo que melhor se ajusta aos dados. Além disso, foram aplicados algoritmos de Aprendizado de Máquina, incluindo *Random Forest*, *Support Vector Machine (SVM)*, *Árvore de Decisão* e *XGBoost*, com o objetivo de melhorar a precisão preditiva. A validação cruzada e o uso de métricas como RMSE indicaram que o modelo Logístico foi o mais apropriado entre os modelos não lineares, enquanto os algoritmos de Aprendizado de Máquina demonstraram maior capacidade preditiva em cenários complexos. Essa abordagem integrada permitiu simular cenários futuros e realizar ajustes em práticas de manejo, como irrigação e fertilização, proporcionando uma base científica para a tomada de decisões na agricultura.

**Palavras-chave:** modelos não lineares; regressão logística; aprendizado de máquina.

## ABSTRACT

This study investigates the use of nonlinear regression models and machine learning algorithms to describe the growth of onion bulbs over time, aiming to optimize agricultural production. Using experimental data three nonlinear growth models were fitted: *Gompertz*, *Von Bertalanffy*, and Logistic, to the dry weight of the onion bulb, with growth time as the explanatory variable. The analysis was conducted using R software, employing selection criteria such as AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) to identify the model that best fits the data. Additionally, machine learning algorithms, including *Random Forest*, *Support Vector Machine (SVM)*, decision tree and *XGBoost*, were applied to improve predictive accuracy. Cross-validation and metrics such as RMSE indicated that the Logistic model was the most appropriate among the nonlinear models, while the machine learning algorithms demonstrated greater predictive capacity in complex scenarios. This integrated approach allows for simulating future scenarios and making adjustments to management practices, such as irrigation and fertilization, providing a scientific basis for decision-making in agriculture.

**Keywords:** nonlinear models; logistic regression; machine learning.

## LISTA DE FIGURAS

4.1	Gráfico de dispersão para o Peso seco e o Tempo de crescimento do bulbo da cebola. ....	29
4.2	Gráfico de ajuste do modelo não linear Gompertz para o Peso seco do bulbo da cebola e Tempo de crescimento. ....	30
4.3	Gráfico de ajuste do modelo não linear Von Bertalanffy para o Peso seco do bulbo da cebola e Tempo de crescimento. . . . .	30
4.4	Gráfico de ajuste do modelo não linear Logístico para o Peso seco do bulbo da cebola e Tempo de crescimento. ....	31
4.5	Normalidade dos Resíduos. ....	32
4.6	Gráfico dos resíduos estudentizados do modelo não linear Logístico. . . . .	33
4.7	Gráfico de compração dos melhores modelos. . . . .	35

## LISTA DE TABELAS

2.1	Modelos de regressão não-linear. . . . .	14
2.2	Modelos de rendimento e suas funções. . . . .	16
4.1	Estatísticas descritivas do Bulbo da Cebola. . . . .	28
4.2	Valores de estimativa dos parâmetros dos modelos não lineares; erro padrão de estimativa, valor- $p$ e intervalo de confiança. . . . .	29
4.3	Critérios AIC e BIC para os modelos não lineares. . . . .	31
4.4	Teste de Shapiro-Wilk.	33
4.5	Comparação de medidas para os modelos de Aprendizado de Máquina. . .	34
4.6	Comparação do RMSE e Acurácia entre Modelo Logístico e Random Forest.	34

# SUMÁRIO

	Página
<b>1</b>	<b>INTRODUÇÃO</b> <b>11</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> <b>13</b>
<b>2.1</b>	<b>Modelos Não Lineares</b> . . . . . 13
2.1.1	Modelos de Crescimento . . . . . 14
2.1.2	Modelos de Rendimento . . . . . 15
<b>2.2</b>	<b>Métodos de Estimação</b> . . . . . 16
2.2.1	Método dos Mínimos Quadrados . . . . . 16
2.2.2	Método dos Mínimos Quadrados Generalizados . . . . . 17
2.2.3	Método da Máxima Verossimilhança . . . . . 18
<b>2.3</b>	<b>Métodos Iterativos</b> . . . . . 18
2.3.1	Medidas de não linearidade . . . . . 19
2.3.2	Técnicas de Diagnóstico . . . . . 21
2.3.3	Critérios de informação de Akaike e Bayesiano de Schwarz . . . . . 23
<b>2.4</b>	<b>Aprendizado de Máquina (<i>Machine Learning</i>)</b> . . . . . 24
2.4.1	Comparação com Modelos Convencionais . . . . . 25
<b>3</b>	<b>MATERIAL E MÉTODOS</b> <b>27</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÕES</b> <b>28</b>
<b>4.1</b>	<b>Modelos Não-Lineares</b> . . . . . 28
<b>4.2</b>	<b>Aprendizado de Máquina</b> . . . . . 33
<b>4.3</b>	<b>Comparação dos modelos</b> . . . . . 34
<b>5</b>	<b>CONCLUSÃO</b> <b>36</b>
	<b>REFERÊNCIAS</b> . . . . . 37

## 1 INTRODUÇÃO

O cultivo da cebola (*Allium cepa* L.) é uma atividade agrícola de grande relevância mundial, contribuindo significativamente para a economia em diversas regiões (Faria et al., 2006). O bulbo da cebola, parte comestível da planta e alvo do interesse comercial, passa por um complexo processo de crescimento, influenciado por fatores ambientais e genéticos. Compreender esse processo é vital para maximizar a produtividade e melhorar a qualidade do produto final, permitindo maior eficiência no manejo agrícola.

Os modelos de crescimento são amplamente aplicados em estudos agrônômicos para descrever a relação entre variáveis temporais e biométricas, como o peso seco de órgãos vegetais ao longo do tempo (Evans, 1972; Guimarães, 1994). Esses modelos possibilitam a previsão da evolução do crescimento, auxiliando na tomada de decisões estratégicas no campo. Contudo, a natureza complexa e não linear do crescimento da cebola demanda o uso de modelos estatísticos capazes de capturar essa complexidade com precisão.

Além dos modelos tradicionais, algoritmos de Aprendizado de Máquina (*Machine Learning*), como afirma Alpaydin (2020), vêm se destacando como ferramentas poderosas para análise de dados agrícolas, oferecendo uma abordagem empírica que pode melhorar a precisão preditiva. Métodos como *Random Forest* (Breiman, 2001), *Support Vector Machine (SVM)* (Cortes; Vapnik, 1995), *Árvore de Decisão* (Quinlan, 1986) e *XGBoost* (Chen; Guestrin, 2016) são eficazes para lidar com a complexidade e a não linearidade dos dados, identificando padrões e interações que poderiam não ser capturados por modelos de crescimento convencionais. A aplicação dessas técnicas pode, assim, complementar as análises tradicionais e resultar em previsões mais robustas sobre o crescimento da cebola.

Modelos não lineares, como o *Gompertz* (Samoletov; Ivanov; Popov, 2017), *Von Bertalanffy* (Von Bertalanffy, 1934) e *Logístico* (Pomeroy; Williams; Moore, 1997), são especialmente úteis para descrever o comportamento assintótico do crescimento, no qual a taxa de crescimento desacelera à medida que a planta atinge a maturidade. Estes modelos são adequados para representar fenômenos biológicos com crescimento inicial rápido, seguido por desaceleração até o ponto de saturação. Medidas de não linearidade, como as desenvolvidas por Beale (1960), Box (1971) e Bates e Watts (1980), desempenham um papel crucial na avaliação desses modelos, quantificando sua complexidade e auxiliando na escolha do modelo mais apropriado.

Nesse contexto, este trabalho propõe aplicar e comparar diferentes modelos de regressão não linear e algoritmos de Aprendizado de Máquina para descrever o crescimento do bulbo da cebola. A análise incluirá medidas de não linearidade e critérios de seleção de modelos, com o objetivo de identificar o modelo mais adequado para descrever a curva de crescimento. Além disso, espera-se que as previsões obtidas auxiliem na tomada de decisões agrícolas, como o planejamento de colheitas e a alocação de recursos. Com essa

abordagem, busca-se contribuir para uma compreensão mais profunda do desenvolvimento da cebola, promovendo práticas agrícolas mais eficientes e sustentáveis.

Além disso, a integração dessas técnicas pode resultar em um avanço significativo no manejo agrícola, permitindo não apenas uma previsão mais precisa do crescimento das plantas, mas também a adaptação contínua das práticas agrícolas às condições variáveis do ambiente. A combinação de métodos tradicionais e modernos, como a modelagem não linear associada a algoritmos de Aprendizado de Máquina, tem o potencial de transformar a forma como os agricultores planejam e executam suas atividades, contribuindo para a produção de cebolas de alta qualidade de forma mais sustentável. Esse impacto se estende além das fronteiras do cultivo da cebola, podendo ser adaptado a outras culturas, ampliando os benefícios das abordagens preditivas para a agricultura de precisão em uma escala global.

## 2 FUNDAMENTAÇÃO TEÓRICA

Nessa seção serão apresentados os métodos e modelos que fundamentaram a pesquisa, assim como os critérios utilizados para sua seleção. Serão detalhadas as técnicas de ajuste dos modelos *Gompertz*, *Von Bertalanffy* e Logístico.

### 2.1 Modelos Não Lineares

Os modelos de regressão são ferramentas estatísticas fundamentais que permitem a análise e a previsão de fenômenos complexos em diversas áreas do conhecimento, incluindo agricultura, biologia, econometria, engenharia e química, segundo Montgomery e Peck (1992), Bates e Watts (1988). Esses modelos ajudam a descrever a relação entre variáveis, possibilitando uma compreensão mais profunda dos dados. Entre os diferentes tipos de modelos de regressão, os modelos não lineares têm se destacado pela sua capacidade de capturar dinâmicas mais complexas que não podem ser adequadamente representadas por relações lineares (Myers, Raymond H., 1990), (Box; Draper, 1987).

Os modelos não lineares são particularmente úteis quando a relação entre as variáveis não é constante e pode variar em diferentes condições. O modelo de regressão não linear com erros normais é um dos métodos mais aplicados, especialmente quando a variável de resposta é contínua. Esse modelo é amplamente adotado devido à sua facilidade de aplicação e à riqueza das teorias desenvolvidas em torno dele.

Compreender os modelos não lineares é essencial para o avanço das análises estatísticas, especialmente na agricultura, no qual fenômenos como o crescimento de plantas podem ser influenciados por múltiplos fatores. Modelos como os de *Gompertz*, *Von Bertalanffy* e Logístico oferecem maneiras eficazes de descrever essas complexidades e são frequentemente utilizados em pesquisas para otimizar a produção agrícola.

A expressão matemática que descreve um modelo de regressão não linear pode ser representada da seguinte maneira:

$$y_i = f(x_i, \theta) + \epsilon_i, \quad i = 1, \dots, n. \quad (2.1)$$

Assim, em um modelo de regressão não-linear, a variável resposta ( $y_i$ ) é modelada como uma função não-linear contínua ( $f(\cdot)$ ) dos parâmetros desconhecidos ( $\theta$ ) e do vetor de variáveis explicativas ( $x_i$ ). Em outras palavras, a variável resposta depende de uma combinação não-linear das variáveis explicativas e dos parâmetros desconhecidos, representada pela função  $f(\cdot)$ , que é conhecida.

De acordo com a explicação de Bates e Watts (1988), nos modelos de regressão não linear, a presença de não linearidade nos parâmetros é evidenciada quando, ao calcularmos as derivadas parciais da função  $f$  em relação a estes parâmetros, percebemos que pelo menos uma delas depende de pelo menos um dos parâmetros do modelo.

### 2.1.1 Modelos de Crescimento

Os modelos de crescimento são essenciais para representar matematicamente o aumento ou desenvolvimento de variáveis ao longo do tempo. No contexto da modelagem não linear, eles se destacam pela capacidade de descrever dinâmicas de crescimento em resposta a fatores variados. Em particular, modelos como os de *Gompertz*, *Von Bertalanffy* e Logístico são amplamente utilizados na pesquisa agrícola, permitindo uma análise detalhada do crescimento de plantas, como o bulbo da cebola.

Esses modelos oferecem uma compreensão aprofundada de como diferentes condições ambientais e práticas de manejo influenciam o desenvolvimento das culturas. Ao integrar essas abordagens com os conceitos de regressão não linear, é possível otimizar a produção agrícola, melhorando as estratégias de cultivo e manejo.

Tabela 2.1 – Modelos de regressão não-linear.

Modelos	Função
Logístico	$f(x) = \frac{\alpha}{1+e^{(\beta-\gamma x)}}$
Gompertz	$f(x) = \alpha e^{-e^{(\beta-\gamma x)}}$
Von Bertalanffy	$f(x) = \alpha[1 - e^{-\gamma(x-\delta)}]$
Richards	$f(x) = \frac{\alpha}{1+e^{(\beta-\gamma x)^{1/\delta}}}$
Weibull	$f(x) = \alpha - \beta e^{-\gamma x^\delta}$
Morgan-Mercer-Flodin (MMF)	$f(x) = \frac{\beta\gamma + \alpha x^\delta}{\gamma + x^\delta}$

Fonte: (Seber; Wild, 2003).

A análise dos dados de crescimento é importante em muitas áreas de pesquisa segundo Seber e Wild (2003). Na Tabela 2.1 são apresentados alguns dos principais modelos de crescimento utilizados para descrever curvas de crescimento.

- **Logístico:** Este modelo é amplamente usado para descrever o crescimento populacional em biologia. A função logística tem formato de curva em S, que começa com um crescimento lento, acelera rapidamente e, depois, desacelera à medida que atinge um limite superior (Bates; Watts, 1988).
- **Gompertz:** Este modelo é utilizado principalmente para descrever o crescimento de tumores e outros sistemas biológicos. A função Gompertz é uma curva em forma de S invertido, com crescimento inicial rápido, seguido de desaceleração e aproximação a um limite inferior (Montgomery; Peck, 1992).
- **Von Bertalanffy:** Modelo comumente empregado para o crescimento de organismos, como plantas e animais. A função Von Bertalanffy começa com um crescimento

acelerado e, posteriormente, desacelera conforme se aproxima de um limite superior (Von Bertalanffy, 1934).

- *Richards*: Um modelo flexível para descrever o crescimento de plantas e animais, também em forma de curva S. Ele ajusta os parâmetros de crescimento mais livremente, acomodando diferentes padrões (Myers, Robert H., 1990).
- *Weibull*: Este modelo é amplamente usado na engenharia para modelar taxas de falhas de sistemas e equipamentos. A função Weibull exhibe um crescimento com uma alta taxa de falhas que diminui com o tempo (Bates; Watts, 1988).
- *Morgan-Mercer-Flodin* (MMF): Um modelo comumente aplicado em estudos biológicos para descrever o crescimento de organismos. A função MMF também segue o formato de curva S, com fases de crescimento lento, rápido e uma desaceleração final (Bates; Watts, 1988).

### 2.1.2 Modelos de Rendimento

Os modelos de crescimento e rendimento estão interligados de forma significativa na busca por otimização agrícola. Enquanto os modelos de crescimento como *Gompertz*, *Von Bertalanffy* e Logístico são utilizados para descrever o desenvolvimento individual da planta ao longo do tempo, os modelos de rendimento focam na relação entre a densidade de plantio e o rendimento total da cultura. A análise conjunta desses modelos ajuda a entender como o crescimento de cada planta afeta o rendimento total da lavoura, permitindo ajustar práticas de manejo para maximizar a produção.

Diversos modelos, que possuem uma forte conexão com o crescimento vegetal, são aplicados para mensurar a ligação entre o rendimento total de uma cultura ( $y$ ) e a densidade de plantação ( $x$ ). À medida que a densidade aumenta, a competição entre as plantas por recursos como água, luz e nutrientes também aumenta, levando a uma queda na produtividade por planta ( $y/x$ ). Dessa forma, o objetivo dos modelos de rendimento é identificar a densidade de plantio que maximize o rendimento total, ao mesmo tempo que permite comparar esses resultados em diferentes cenários e condições.

Na Tabela 2.2, são apresentados alguns dos modelos de rendimento mais utilizados, os quais ajudam os especialistas agrícolas a definir práticas otimizadas de plantio que podem resultar em uma produção mais eficiente e sustentável.

Tabela 2.2 – Modelos de rendimento e suas funções.

Modelos	Função
Holliday et al. (1960)	$f(x) = \frac{x}{\alpha + \beta x + \gamma x^2}$
Bleasdale e Nelder (1960)	$f(x) = x(\alpha + \beta x \phi)^{-1/\phi}$
Ricker (1954)	$f(x) = \alpha x e^{-x/k}$
Hiperbólico	$f(x) = \frac{1}{\alpha + \beta x}$

Fonte: (Holliday; Adams; Anderson, 1960), (Bleasdale; Nelder, 1960), (Ricker, 1954).

O Modelo de Holliday, Adams e Anderson (1960) considera a influência da densidade de plantio nos fatores  $(\alpha)$ ,  $(\beta)$  e  $(\gamma)$ . No modelo de Bleasdale e Nelder (1960),  $(\phi)$  e  $(\theta)$  desempenham papéis importantes na relação entre densidade e rendimento. Já o modelo de Ricker (1954) incorpora uma taxa de mortalidade das plantas em função da densidade. E o modelo Hiperbólico expressa a relação inversa entre densidade e rendimento.

## 2.2 Métodos de Estimação

Existem várias abordagens para obter as estimativas dos parâmetros, incluindo o uso do método dos mínimos quadrados ou do método da máxima verossimilhança. No caso de modelos de regressão não lineares, resolver o sistema de equações normais não é uma tarefa simples, exigindo o emprego de métodos iterativos para alcançar essas estimativas.

### 2.2.1 Método dos Mínimos Quadrados

O método dos mínimos quadrados é uma ferramenta crucial na análise estatística, frequentemente utilizada para estimar os parâmetros de modelos matemáticos. Esse método desempenha um papel essencial ao ajustar modelos aos dados observados, garantindo uma representação adequada do fenômeno em estudo.

Ao considerarmos um modelo representado por uma função  $f(\mathbf{x}_i, \boldsymbol{\theta})$ , em que  $\mathbf{x}_i$  são os valores observados e  $\boldsymbol{\theta}$  são os parâmetros a serem estimados, os mínimos quadrados buscam encontrar uma estimativa  $\hat{\boldsymbol{\theta}}$  que minimize a soma dos quadrados dos resíduos. Esses resíduos são as diferenças entre os valores observados  $y_i$  e os valores previstos pelo modelo  $f(\mathbf{x}_i, \boldsymbol{\theta})$ .

Matematicamente, a soma dos quadrados dos resíduos  $SQRes(\boldsymbol{\theta})$  é dada por:

$$SQRes(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(\mathbf{x}_i, \boldsymbol{\theta})]^2. \quad (2.2)$$

Para encontrar o estimador de mínimos quadrados, é necessário diferenciar a  $SQRes(\boldsymbol{\theta})$

com relação a cada um dos parâmetros e igualar as equações a zero, isto é:

$$\frac{\partial SQRes(\boldsymbol{\theta})}{\partial \theta_r} = 0 \quad \text{para } r = 1, 2, \dots, p. \quad (2.3)$$

Além disso, tem-se a condição:

$$\sum_{i=1}^n (y_i - f(x_i, \boldsymbol{\theta})) \frac{\partial SQRes(\boldsymbol{\theta})}{\partial \theta_r} = 0 \quad \text{para } r = 1, 2, \dots, p. \quad (2.4)$$

Essas equações resultam em um conjunto de equações normais que descrevem o sistema. No entanto, em muitos casos, é impossível resolver essas equações de forma analítica. Portanto, são utilizados métodos iterativos para encontrar soluções aproximadas, ajustando gradualmente os parâmetros do modelo até que as equações se aproximem o suficiente da solução.

Esses métodos iterativos são especialmente cruciais para estimar os parâmetros de modelos não lineares, uma vez que não há soluções analíticas diretas disponíveis. A aplicação eficaz do método dos mínimos quadrados não só permite uma melhor compreensão dos dados, mas também desempenha um papel vital em uma ampla gama de campos, incluindo ciências naturais, engenharia e ciências sociais.

## 2.2.2 Método dos Mínimos Quadrados Generalizados

O Método dos Mínimos Quadrados Generalizados é empregado em situações nas quais os erros apresentam autocorrelação, conforme descrito por Gallant (1987). Nesse método, a função a ser minimizada é a soma dos quadrados dos resíduos, que representa a discrepância entre os valores observados e os valores preditos pelo modelo.

A expressão para a soma dos quadrados dos resíduos é dada por:

$$SQRes(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(\mathbf{x}_i, \boldsymbol{\theta})]^T \mathbf{V}^{-1} [y_i - f(\mathbf{x}_i, \boldsymbol{\theta})], \quad (2.5)$$

Seja  $\mathbf{b}(\hat{\boldsymbol{\theta}}_G)$  a estimativa de mínimos quadrados generalizados que minimiza essa expressão. Suponha que tenhamos uma decomposição de *Cholesky* de  $\mathbf{V}$  dada por  $\mathbf{V}' = \mathbf{U}^T \mathbf{T}$ , em que  $\mathbf{U}$  é uma matriz triangular superior. Ao multiplicarmos o modelo não linear por  $\mathbf{R} = (\mathbf{U}^T)^{-1}$ , obtem-se:

$$\mathbf{z} = \mathbf{k}(\boldsymbol{\theta}) + \boldsymbol{\eta}, \quad (2.6)$$

em que,  $\mathbf{z} = \mathbf{R}\mathbf{y}$ ,  $\mathbf{k}(\boldsymbol{\theta}) = \mathbf{R}f(\mathbf{x}, \boldsymbol{\theta})$  e  $\boldsymbol{\eta} = \mathbf{R}$ . Assim, a soma de quadrados dos resíduos é dada por:

$$SQRes(\boldsymbol{\theta}) = [\mathbf{z} - \mathbf{k}(\boldsymbol{\theta})]^T [\mathbf{z} - \mathbf{k}(\boldsymbol{\theta})]. \quad (2.7)$$

A partir desse momento, o vetor de estimativas  $\hat{\theta}_G$  passa a ser um vetor de estimativas de mínimos quadrados para os dados transformados, pois essa transformação utilizando a decomposição de *Cholesky* permite tratar os dados como um caso de variâncias homogêneas com observações independentes.

### 2.2.3 Método da Máxima Verossimilhança

O Método da Máxima Verossimilhança é uma técnica estatística para estimar os parâmetros de um modelo, maximizando a função de verossimilhança. Esta função mede a probabilidade dos dados observados sob diferentes valores dos parâmetros do modelo. No contexto de modelos não lineares com erros normalmente distribuídos, a função de verossimilhança é expressa como um produto de densidades normais, cada uma dependendo da diferença entre os dados observados e os valores ajustados pelo modelo. A maximização da função de verossimilhança permite encontrar as estimativas dos parâmetros que mais provavelmente gerariam os dados observados. No caso de modelos não lineares, a estimativa de máxima verossimilhança é particularmente útil, pois muitos fenômenos naturais são melhor descritos por modelos que capturam relações não lineares entre variáveis (Casella; Berger, 2002).

A função de verossimilhança pode ser expressa como:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\mathbf{y}_i - f(\mathbf{x}_i, \boldsymbol{\theta}))^2}{2\sigma^2}\right), \quad (2.8)$$

em que,  $L(\boldsymbol{\theta})$  é a função de verossimilhança,  $\mathbf{y}_i$  são os dados observados,  $f(\mathbf{x}_i, \boldsymbol{\theta})$  são os valores ajustados pelo modelo,  $\boldsymbol{\theta}$  representa os parâmetros do modelo e  $\sigma^2$  é a variância dos erros normalmente distribuídos.

## 2.3 Métodos Iterativos

Métodos iterativos são fundamentais em diversas áreas da ciência e da engenharia, especialmente na estimação de parâmetros em modelos não lineares que não existe uma solução analítica direta. Entre os métodos iterativos mais amplamente utilizados, destaca-se o método de *Gauss-Newton*. Este método é particularmente eficaz na estimação de parâmetros de modelos de regressão não linear, conforme citam Bates e Watts (1988) e Seber e Wild (2003).

O método de *Gauss-Newton* busca minimizar a soma dos quadrados dos resíduos entre os valores observados e os preditos pelo modelo. A função a ser minimizada é expressa por:

$$SQRes(\boldsymbol{\theta}) = \sum_{i=1}^n (\mathbf{y}_i - f(\mathbf{x}_i, \boldsymbol{\theta}))^2, \quad (2.9)$$

em que  $\mathbf{y}_i$  são os dados observados,  $f(\mathbf{x}_i, \boldsymbol{\theta})$  são os valores preditos pelo modelo, e  $\boldsymbol{\theta}$  representa os parâmetros do modelo.

A técnica se baseia na aproximação da função de resposta por uma função linear através de iterações. A cada iteração, calcula-se a matriz jacobiana da função de resposta em relação aos parâmetros e, em seguida, utiliza-se essa matriz para atualizar os parâmetros do modelo.

O processo iterativo continua até que uma condição de convergência seja alcançada, como quando a mudança nos parâmetros estimados entre duas iterações consecutivas se torna suficientemente pequena. O método de *Gauss-Newton* é especialmente eficiente quando o modelo é bem comportado e próximo da solução, convergindo rapidamente para o mínimo local da função objetivo (Nocedal; Wright, 2006).

Além do método de *Gauss-Newton*, outros métodos iterativos comumente utilizados incluem o método de *Newton-Raphson*, o método do gradiente descendente (ou método do maior declive), e o método de *Levenberg-Marquardt*. Cada um desses métodos possui características próprias e se adequa a diferentes tipos de problemas (Fletcher, 1987) e (Gill; Murray; Wright, 1981).

Em suma, os métodos iterativos desempenham um papel crucial na estimação de parâmetros em modelos não lineares, permitindo soluções aproximadas em situações no qual não é possível encontrar uma solução analítica direta. O método de *Gauss-Newton* se destaca por sua eficiência e ampla aplicabilidade em diversos contextos de modelagem.

### 2.3.1 Medidas de não linearidade

Diversos autores desenvolveram medidas para avaliar a não linearidade em problemas de estimação não linear. Beale (1960) propôs medidas iniciais, embora estas tendam a subestimar a não linearidade. Box (1971) introduziu uma medida de viés para estimadores de máxima verossimilhança em modelos não lineares, que foi validada por Gillis e Ratkowsky (1978). Bates e Watts (1980) criaram medidas de curvatura que indicam a não linearidade intrínseca e aparente em modelos normais não lineares. Na prática, as medidas de Box e de curvatura de Bates e Watts são as mais utilizadas para avaliar a não linearidade, sendo úteis não apenas em estimação, mas também em problemas de discriminação, em que modelos mais lineares são preferidos.

Além das medidas de Box e da curvatura de Bates e Watts, diversas outras abordagens foram desenvolvidas para quantificar a não linearidade em modelos não lineares. Essas medidas são fundamentais para avaliar se as técnicas tradicionais de estimação, que foram inicialmente concebidas para regressões lineares, são adequadas para uma boa aproximação do modelo não linear em análise. Compreender o grau de não linearidade é crucial, pois isso orienta a escolha da abordagem de modelagem mais apropriada e assegura resultados precisos na análise estatística.

## Medidas de curvatura de Bates e Watts

Bates e Watts (1980) desenvolveram medidas de não linearidade baseadas em conceitos de geometria diferencial. Essas medidas estendem as ideias propostas por Beale (1960) e se fundamentam no conceito geométrico de curvatura. A abordagem deles envolve a quantificação da não linearidade em duas componentes: a primeira refere-se à avaliação da estrutura da função de resposta, enquanto a segunda considera a relação entre a variabilidade dos dados e dos parâmetros do modelo. Para obter tais componentes, inicialmente, a partir da teoria de modelos lineares, é possível estabelecer um intervalo de confiança para o vetor de parâmetros  $\boldsymbol{\theta}$ . Esse intervalo é definido pela desigualdade:

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' X' X (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq ps^2 F_{p,n-p}^\alpha. \quad (2.10)$$

Nessa expressão,  $X$  representa a matriz do modelo e  $\boldsymbol{\theta}$  é o vetor de parâmetros. Os autores também introduziram o raio padrão  $\rho = s\sqrt{p}$ , que corresponde à esfera cujos valores internos pertencem ao intervalo de confiança  $1 - \alpha$ . Essa definição é essencial para garantir que a estimativa dos parâmetros esteja dentro da região de confiança estabelecida.

Para avaliar a validade da aproximação linear, Bates e Watts consideraram a magnitude do segundo termo da expansão de série de Taylor em relação ao primeiro termo. Eles dividiram o componente, a seguir, em dois termos:

$$\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})' H \boldsymbol{\theta}^{(a)} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}).$$

Um pertencente ao plano da aproximação e outro ao espaço complementar, obtido por projeção ortogonal. Essa decomposição é realizada por meio da matriz de projeção:

$$\hat{\boldsymbol{P}} = \hat{\boldsymbol{F}} (\hat{\boldsymbol{F}}' \hat{\boldsymbol{F}})^{-1} \hat{\boldsymbol{F}}'. \quad (2.11)$$

As duas componentes resultantes são:

$$\hat{\boldsymbol{H}} = \hat{\boldsymbol{H}}^T + \hat{\boldsymbol{H}}^N, \quad (2.12)$$

em que,  $T$  e  $N$  representam as componentes tangencial e normal ao plano definido pela aproximação linear, respectivamente. As medidas de curvatura propostas por Bates e Watts comparam o termo quadrático com o termo linear:

$$K^T = \frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \boldsymbol{H}_T (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})}{F(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^2} \quad \text{e} \quad K^N = \frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \boldsymbol{H}_N (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})}{F(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^2}. \quad (2.13)$$

Essas curvaturas correspondem aos círculos que melhor aproximam o formato da superfície esperada em um modelo  $\eta(\hat{\boldsymbol{\theta}})$ . Os círculos têm raio  $R = 1/K$ , e para obter medidas invariantes à escala da variável resposta, os autores padronizaram essas medidas pelo raio

padrão  $\rho$ , tais que:

$$\Gamma^T = K^T \rho \quad \text{e} \quad \Gamma^N = K^N \rho. \quad (2.14)$$

Nesse sentido, essas medidas são utilizadas para avaliar a não linearidade dos componentes do modelo.  $\Gamma^T$  está associada a desvios de uniformidade e paralelismo do plano tangente correspondente à parametrização do modelo, enquanto  $\Gamma^N$  está associada a desvios da planicidade da superfície esperada e é invariante à reparametrização. Além disso, Bates e Watts sugerem o critério:

$$\Gamma > 2(F_{p,n-p}^\alpha)^{-1/2}. \quad (2.15)$$

Em que o mesmo serve como guia para indicar se o modelo ajustado possui uma curvatura aparentemente acentuada. Vale ressaltar que essas medidas são obtidas para uma direção específica no espaço, e um processo iterativo é necessário para encontrar a maior curvatura. Além disso, há uma relação entre essa medida e os trabalhos anteriores de Beale (1960) e Box (1971). Ratkowsky (1983) apresentou um algoritmo para calcular a máxima curvatura, fornecendo assim uma ferramenta prática para aplicar esses conceitos.

### 2.3.2 Técnicas de Diagnóstico

As Técnicas de Diagnóstico servem para validar modelos estatísticos, verificando se são satisfeitas as suposições envolvidas e avaliando tanto a qualidade quanto a performance dos ajustes do modelo escolhido. Entre os modelos de diagnóstico, destacam-se a Matriz de Projeção, Resíduo Projetado e Medidas de Influência.

#### **Matriz de projeção**

A Matriz de Projeção é utilizada em modelos normais não lineares com o intuito de detectar os pontos que estão mais afastados dos demais. Ela é definida por:

$$\hat{\mathbf{P}} = \mathbf{F}(\hat{\boldsymbol{\theta}})[\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{F}(\hat{\boldsymbol{\theta}})]^{-1}\mathbf{F}'(\hat{\boldsymbol{\theta}}). \quad (2.16)$$

Os pontos afastados, que também atuam como alavancas e influentes, costumam ter um impacto significativo na matriz de variância e covariância Mattos (2013). Essa matriz é de projeção local, pois depende de  $\theta$ , sendo estimada em  $\hat{\theta}$ . Além disso, o critério  $\hat{p}_{ii} \geq 2p/n$  ainda deve ser utilizado como referência na detecção desses pontos influentes.

#### **Resíduo projetado**

Em alguns casos, os resíduos ordinários no modelo produzem resultados enganosos, por exemplo em situações nas quais são utilizados em métodos de diagnósticos análogos aos da regressão linear, e, portanto, não refletindo corretamente a distribuição dos erros.

Cook e Tsai (1985) apresentaram um novo tipo de resíduo a fim de superar deficiências desses resíduos.

Primeiramente, considera-se  $C(\mathbf{F})$  como o subespaço gerado pelas colunas de  $F$  e empregando  $\overline{C}(\mathbf{F})$  como representação do espaço ortogonal a  $C(\mathbf{F})$ . Em seguida, são definidas as matrizes  $p \times p$ :

$$\mathbf{W}_i = \left( \frac{\partial^2 f(x_i, \boldsymbol{\theta})}{\partial \theta_r \partial \theta_s} \right), \quad r, s = 1, \dots, p,$$

sendo estas, derivadas de segunda ordem para  $i = 1, \dots, n$ . Além disso,  $\mathbf{W}$  é uma matriz  $n \times (p \times p)$ , no qual a  $i$ -ésima face é igual a  $\mathbf{W}_i$ . Os autores supracitados ainda apresentam uma aproximação para o resíduo ordinário  $r$ , tal que:

$$\mathbf{r} \cong (\mathbf{I} - \mathbf{P})\boldsymbol{\epsilon} - \mathbf{F} \sum_{i=1}^n r_i \mathbf{W}_i \boldsymbol{\delta} - \frac{1}{2} (\mathbf{I} - \mathbf{P}) \boldsymbol{\delta}' \mathbf{W} \boldsymbol{\delta}, \quad (2.17)$$

sendo que  $\boldsymbol{\delta} = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ , e  $\boldsymbol{\delta}' \mathbf{W} \boldsymbol{\delta}$  representa  $\boldsymbol{\delta}' \mathbf{W}_i \boldsymbol{\delta}$  para produzir a  $i$ -ésima componente de  $\mathbf{r}$  (Mattos, 2013).

Os autores ainda explicam que a matriz  $\mathbf{W}$  apresenta  $p$  faces, cada uma delas uma matriz  $n \times p$ , considerando  $n$  como o número de observações e  $p$  o número de parâmetros. Em seguida, cada coluna de derivadas de segunda ordem da matriz  $\mathbf{W}$  é projetada sobre  $\overline{C}(\mathbf{F})$ , gerando-se o operador de projeção ortogonal  $\mathbf{P}_1 = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$ . Deve ser informado que  $\mathbf{S} = (\mathbf{I} - \mathbf{P})\mathbf{T}$ , em que  $\mathbf{T}$  é uma matriz  $n \times q$  obtida pelos vetores não nulos, de derivadas de segunda ordem e  $\mathbf{P} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$ .

Há ainda a matriz  $\mathbf{V}$  de ordem  $n \times (p+q)$ , sendo definida como  $\mathbf{V} = (\mathbf{F}, \mathbf{S})$  e o operador de projeção ortogonal em  $C(\mathbf{V})$ ,  $\mathbf{P}_2 = \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'$ . Todas as matrizes esclarecidas anteriormente foram utilizadas por Cook e Tsai (1985) para definir o resíduo projetado, dado por

$$(\mathbf{I} - \mathbf{P}_2)\mathbf{r} = (\mathbf{I} - \mathbf{P})\boldsymbol{\epsilon} - (\mathbf{I} - \mathbf{P}_1)\boldsymbol{\epsilon}.$$

De modo geral, os resíduos projetados oferecem uma qualidade diagnóstica superior em comparação aos resíduos ordinários. Além disso, muitas das técnicas de diagnóstico empregadas na regressão linear podem ser aplicadas a esses resíduos projetados.

Entretanto, vale lembrar que a regressão linear apresenta resíduos correlacionados e variâncias diferentes, mesmo em casos em que os erros não são correlacionados e de variância constante. Pensando-se nisso, foram definidos resíduos estudentizados, esses com média zero e variância constante igual a 1, ainda que sejam correlacionados tais resíduos. Expressa-se o  $i$ -ésimo resíduo ordinário estudentizado da seguinte forma:

$$t_i = \frac{r_i}{s(1 - p_{ii})^{\frac{1}{2}}}, \quad i = 1, \dots, n. \quad (2.18)$$

Deve-se ainda dizer que os os resíduos projetados estudentizados têm esperança nula

e variância igual a  $\sigma^2 = (\mathbf{I} - \mathbf{P}_2)$ , sendo assim definidos:

$$s_i = \frac{\{(\mathbf{I} - \hat{\mathbf{P}}_2)\mathbf{r}\}_i}{\tilde{\sigma}\{(\mathbf{I} - \hat{\mathbf{P}}_2)\mathbf{r}\}_{ii}^{1/2}}. \quad (2.19)$$

### Medidas de Influência

Na Análise de Diagnóstico, é essencial reconhecer observações que podem exercer uma influência significativa nas estimativas dos parâmetros do modelo. A presença desses pontos atípicos pode alterar de forma substancial a interpretação dos resultados, comprometendo a validade das conclusões tiradas a partir do modelo estatístico. De acordo com Mattos (2013), as medidas de influência para modelos normais não lineares são baseadas na regressão linear. Para esses modelos, a estimativa dos parâmetros sem a  $i$ -ésima observação, denotada como  $\hat{\boldsymbol{\theta}}_{(i)}$ , é obtida inicializando o processo iterativo a partir da estimativa  $\hat{\boldsymbol{\theta}}$  e tomando uma estimativa de um passo. Essa abordagem é descrita por Cook e Weisberg (1982) e é fundamental para avaliar a influência de observações específicas no modelo.

$$\hat{\boldsymbol{\theta}}_{(i)} = \hat{\boldsymbol{\theta}} - \frac{(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}}{1 - \hat{p}_{ii}} \mathbf{f}_i r_i. \quad (2.20)$$

Outra medida de influência bastante utilizada é a Distância de *Cook*, dada por:

$$D_i = (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)})' (\hat{\mathbf{F}}'\hat{\mathbf{F}}) (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)}) / p\hat{\sigma}^2. \quad (2.21)$$

#### 2.3.3 Critérios de informação de Akaike e Bayesiano de Schwarz

Conforme mencionado por Ueda (2003), há vários critérios disponíveis para a comparação e seleção de modelos que apresentam diferentes quantidades de parâmetros. Entre esses critérios, destacam-se o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC), que são utilizados para determinar qual modelo se ajusta de forma mais adequada a um conjunto de dados específico ou que apresenta um maior poder explicativo. O AIC é definido como:

$$\text{AIC} = 2K - 2\ln(\hat{L}), \quad (2.22)$$

enquanto o BIC, por sua vez, é definido como:

$$\text{BIC} = -2\log(\hat{L}) + \log(N)K, \quad (2.23)$$

sendo  $\hat{L}$  a Máxima Verossimilhança,  $K$  o número de parâmetros a ser estimados no modelo e  $N$  o tamanho da amostra.

Ao analisar um modelo por esses critérios, deve-se escolher aquele que apresentar menores valores de AIC ou BIC. É importante mencionar que autores como Burnham e

Anderson (2002) concluíram, através de bases teóricas e simulações, que o BIC é o critério de informação mais adequado caso “a complexidade do modelo verdadeiro não cresça com o tamanho do conjunto de dados” (Ueda, 2003). Além disso, segundo as diretrizes estabelecidas por Burnham e Anderson (2002), a diferença entre dois valores de AIC deve ser interpretada com cautela; uma diferença menor que um sugere que os modelos são comparáveis em termos de adequação aos dados. Portanto, não se pode considerar essa diferença como significativa para a escolha entre os modelos.

## 2.4 Aprendizado de Máquina (*Machine Learning*)

O termo Aprendizado de Máquina refere-se à identificação automática de padrões significativos em grandes volumes de dados (Shalev-Shwartz; Ben-David, 2011). Essa técnica é amplamente utilizada em tarefas como reconhecimento, diagnóstico, planejamento e predição, frequentemente associadas à inteligência artificial (IA) (Nilsson, 1998). Os métodos de *Machine Learning* combinam estatísticas e otimização matemática para inferir modelos preditivos a partir dos dados, aplicando conjuntos de treino e teste para ajustar e validar suas previsões (Daniel, 2021).

Neste estudo, algoritmos de Aprendizado de Máquina foram empregados para modelar o crescimento do bulbo da cebola, que envolve relações não-lineares complexas. Enquanto os modelos convencionais, como *Gompertz*, *Von Bertalanffy* e Logístico, são baseados em suposições teóricas, os algoritmos de Aprendizado de Máquina oferecem uma abordagem empírica, capaz de lidar com interações complexas entre as variáveis (Ryo; Rillig, 2017).

O *Random Forest* é um dos algoritmos aplicados neste trabalho, conhecido por criar diversas árvores de decisão e utilizar a média dos resultados para gerar previsões robustas. Esse método também permite a avaliação da importância dos preditores com base em permutações, o que facilita a identificação dos fatores mais relevantes (Ryo; Rillig, 2017). Além disso, o *Support Vector Machine (SVM)* será utilizado, um método que transforma os dados de entrada em um espaço de características de alta dimensão e encontra o hiperplano que separa as classes de forma ótima. O SVM é útil tanto para tarefas de classificação quanto de regressão, devido à sua robustez contra ruídos e capacidade de evitar o sobreajuste (Awad; Khanna, 2015).

Outro algoritmo relevante aplicado será o *XGBoost*, um método de *boosting* que ajusta sequencialmente modelos simples (*base learners*) para corrigir os erros cometidos pelos modelos anteriores. O XGBoost é particularmente eficaz para ajustar relações não-lineares complexas e lida bem com dados que apresentam interações entre variáveis (Ding; Cao; Næss, 2021). Cada um desses algoritmos será ajustado e validado utilizando técnicas apropriadas de *Machine Learning*, como a validação cruzada, para garantir que os modelos serão generalizáveis a novos dados.

Além dos métodos mencionados, o modelo de Árvore de Decisão também será considerado em nossa análise. As árvores de decisão são algoritmos que utilizam uma estrutura de

árvore para tomar decisões baseadas em regras simples, resultantes da divisão dos dados em subconjuntos homogêneos. Cada nó interno da árvore representa uma pergunta sobre uma característica, e as folhas finais representam as previsões. Este método é intuitivo e fácil de interpretar, o que facilita a visualização das decisões tomadas pelo modelo.

Nesse sentido, os algoritmos de Aprendizado de Máquina requerem a escolha cuidadosa de hiperparâmetros para otimizar o desempenho dos modelos. No caso do *Random Forest*, isso inclui a definição do número de árvores geradas e da profundidade máxima das árvores. Para o SVM, a escolha do *kernel* e o valor do parâmetro de regularização são fundamentais para o ajuste do modelo. Já no XGBoost, a taxa de aprendizado e o número de iterações são parâmetros críticos para evitar tanto o subajuste quanto o sobreajuste. O ajuste desses hiperparâmetros será realizado utilizando técnicas de validação cruzada, o que permite garantir que os modelos serão robustos e tenham uma boa capacidade preditiva, sem superestimar o desempenho nos dados de treino. Esses ajustes são essenciais para equilibrar a complexidade do modelo e a precisão preditiva.

#### 2.4.1 Comparação com Modelos Convencionais

Os modelos convencionais, como *Gompertz*, *Von Bertalanffy* e Logístico, são amplamente utilizados para descrever o crescimento biológico, baseando-se em funções teóricas que possibilitam uma interpretação clara dos parâmetros. Esses modelos representam uma abordagem clássica para a análise do crescimento, partindo de premissas sobre a forma funcional e com estimativas de causalidade geradas de maneira tradicional, fundamentadas em critérios estatísticos.

Entretanto, esses modelos podem apresentar limitações ao lidar com dados complexos e não lineares, como os relacionados ao crescimento do bulbo da cebola. Em contrapartida, os algoritmos de Aprendizado de Máquina, como *Random Forest*, *SVM*, *Árvore de Decisão* e *XGBoost*, oferecem maior flexibilidade para capturar padrões não-lineares e interações complexas entre variáveis. Embora esses algoritmos exijam mais recursos computacionais e sejam menos interpretáveis em termos de causalidade direta, eles têm o potencial de fornecer previsões mais precisas e confiáveis em cenários de alta complexidade.

A comparação entre modelos convencionais e algoritmos de Aprendizado de Máquina revela um equilíbrio importante: enquanto os primeiros oferecem interpretabilidade dos parâmetros biológicos, os algoritmos de Aprendizado de Máquina destacam-se pela capacidade de adaptação aos detalhes dos dados. No caso do crescimento do bulbo da cebola, esses algoritmos mostram maior precisão nas previsões por sua flexibilidade em capturar padrões não lineares, embora com interpretação mais complexa.

Por fim, ao considerar a aplicação desses modelos, é possível observar que os modelos convencionais são mais indicados quando a compreensão dos parâmetros biológicos é essencial, enquanto os algoritmos de Aprendizado de Máquina se destacam em contextos

em que a precisão preditiva é prioritária, especialmente em conjuntos de dados complexos. Esse contraste justifica uma análise comparativa de ambos os tipos de modelos, permitindo a identificação do mais adequado para diferentes cenários de modelagem de crescimento.

### 3 MATERIAL E MÉTODOS

Os dados utilizados nesse estudo foram coletados de medições do peso seco do bulbo da cebola, considerando o tempo de crescimento como a variável independente. Esses dados foram obtidos a partir do banco de dados do National Institute of Standards and Technology (NIST) (2024). Dada a complexidade da relação entre essas variáveis, serão explorados modelos não lineares para capturar essa dinâmica. Para essa análise, serão avaliados três modelos distintos: o modelo Logístico, o modelo de *Gompertz* e o modelo de *Von Bertalanffy*. Cada modelo será ajustado aos dados para determinar sua capacidade de descrever a relação observada entre o tempo de crescimento e o peso seco do bulbo da cebola.

O ajuste dos modelos será realizado por meio de técnicas apropriadas de ajuste não linear, utilizando o *software* R (Team, 2024), visando identificar os parâmetros que melhor se ajustam aos dados observados. Essa etapa é crucial para selecionar o modelo que represente de forma precisa a relação entre o tempo de crescimento e o peso seco do bulbo da cebola.

Após o ajuste, os resultados serão analisados com base nos parâmetros estimados e nas estatísticas de ajuste. Para cada modelo, serão examinados os valores estimados dos parâmetros, além de medidas de adequação como o Coeficiente de determinação ( $R^2$ ) e os critérios de informação AIC e BIC.

A normalidade dos resíduos de cada modelo ajustado será avaliada utilizando o teste de *Shapiro-Wilk*, que verifica se os resíduos seguem uma distribuição normal. Caso o p-valor associado seja inferior ao nível de significância adotado, rejeitaremos a hipótese nula de normalidade. Neste trabalho, será adotado o nível  $\alpha$  para cada teste. Nesse sentido, os modelos serão comparados com base em medidas de adequação, como AIC e BIC, oferecendo uma abordagem objetiva para a seleção do modelo mais adequado na descrição da relação entre o tempo de crescimento e o peso seco do bulbo da cebola.

Para complementar a análise com modelos não lineares, serão aplicados algoritmos de *Machine Learning* com o objetivo de aprimorar a precisão preditiva da relação entre o tempo de crescimento e o peso seco do bulbo da cebola. Os métodos avaliados incluirão *Random Forest*, *Support Vector Machine (SVM)*, *Árvore de Decisão* e *XGBoost*, escolhidos por sua capacidade de capturar padrões complexos e não lineares nos dados.

O processo de ajuste dos modelos envolverá a divisão dos dados em conjuntos de treino e teste, além da aplicação de técnicas de validação cruzada para otimizar hiperparâmetros específicos de cada modelo. Métricas como o erro quadrático médio (RMSE) serão utilizadas para avaliar o desempenho preditivo desses algoritmos em comparação com modelos convencionais. Essa abordagem robusta proporciona uma compreensão mais aprofundada da dinâmica do crescimento do bulbo da cebola.

## 4 RESULTADOS E DISCUSSÕES

As estatísticas descritivas dos dados são apresentadas na Tabela 4.1. Essa análise fornece uma visão geral das principais características dos dados, permitindo compreender melhor a variabilidade e os padrões presentes na produção dessa cultura. A tabela resume as informações relevantes sobre os parâmetros avaliados.

Tabela 4.1 – Estatísticas descritivas do Bulbo da Cebola.

Medida	Peso Seco (g)	Tempo de Crescimento (dias)
Mínimo	16,08	1.0
1º Quartil (Q1)	144,38	4.5
Mediana	520,53	8.0
Média	423,30	8.0
3º Quartil (Q3)	670,94	11.5
Máximo	724,93	15.0

Fonte: Elaborada pela autora, (2024).

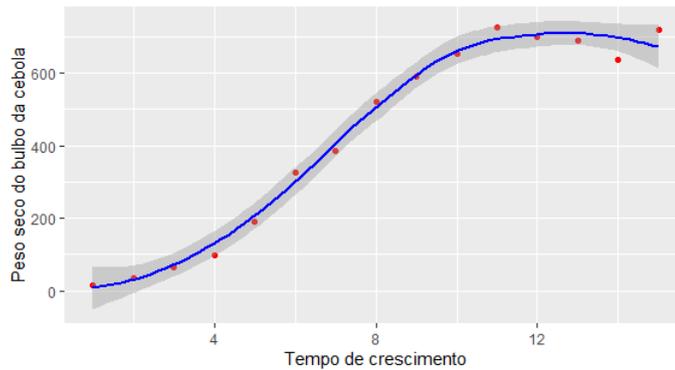
De acordo com a Tabela 4.1, o peso seco dos bulbos varia significativamente, com um mínimo de 16,08 g e um máximo de 724,93 g. A mediana de 520,53 g indica que metade dos bulbos pesou menos que esse valor, sugerindo que muitos bulbos atingem um tamanho considerável. No entanto, a média de 423,30 g é inferior à mediana, indicando a presença de alguns bulbos muito leves, que podem ter se desenvolvido de forma inadequada. Essa diferença entre a média e a mediana pode ser um sinal de que há fatores limitantes que afetam o crescimento, como manejo inadequado, condições climáticas desfavoráveis ou a escolha de variedades menos produtivas.

Em relação ao tempo de crescimento, observa-se que a mediana e a média estão em 8 dias, o que sugere que a maioria das cebolas atinge um desenvolvimento adequado nesse período. O menor tempo registrado foi de apenas 1 dia, o que pode indicar um erro na coleta de dados ou cebolas que não se desenvolveram adequadamente. O máximo de 15 dias sugere que algumas cebolas levaram mais tempo para atingir um tamanho adequado, possivelmente devido a fatores como doenças, pragas ou deficiências nutricionais.

### 4.1 Modelos Não-Lineares

Para modelar o crescimento do peso seco do bulbo da cebola em função do tempo de crescimento, foi aplicado um modelo não linear, adequado para descrever processos biológicos que apresentam crescimento acelerado seguido por uma estabilização. Modelos como *Gompertz*, *Logístico* e *Von Bertalanffy* são comumente utilizados nessas situações, pois capturam bem a dinâmica de crescimento limitada por fatores naturais.

Figura 4.1 – Gráfico de dispersão para o Peso seco e o Tempo de crescimento do bulbo da cebola.



Fonte: Elaborada pela autora, (2024).

Na Figura 4.1, o gráfico de dispersão mostra a relação entre as variáveis mencionadas. Cada ponto representa uma observação experimental, e a linha suavizada reflete o ajuste do modelo não linear escolhido. Observa-se que o crescimento inicial do bulbo é lento, seguido por uma fase de rápido crescimento, até que se estabiliza, atingindo um limite.

Nesse sentido, foram aplicados os modelos não lineares, conforme descrito na metodologia, para ajustar os dados de crescimento do peso seco do bulbo da cebola em função do tempo de crescimento. Os modelos *Gompertz*, Logístico e *Von Bertalanffy* foram escolhidos por sua capacidade de descrever processos biológicos com fases de crescimento inicial lento, seguidas por crescimento acelerado e estabilização.

Na Tabela 4.2 são apresentados os valores calculados para as estimativas dos parâmetros de cada modelo, juntamente com seus respectivos erros padrão, valores de  $t$  e intervalos de confiança a 95%.

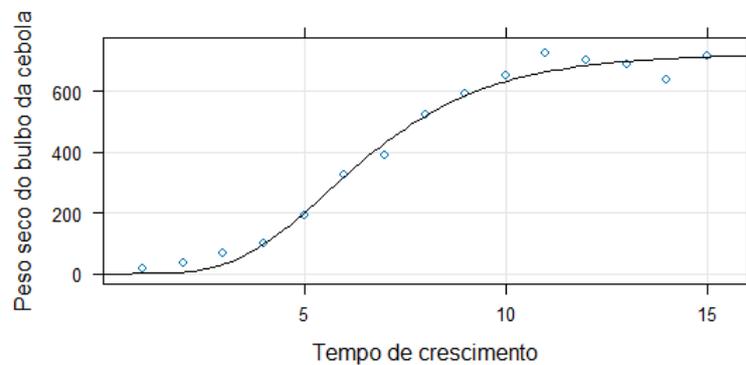
Tabela 4.2 – Valores de estimativa dos parâmetros dos modelos não lineares; erro padrão de estimativa, valor- $p$  e intervalo de confiança.

Modelo Não Linear	Parâmetro	Estimativa	Erro Padrão	Pr(> t )	Intervalo de Confiança(95%)
<i>Gompertz</i>	$\alpha$	723,1087	22,0605	<0,01 ***	[680,19; 775,98]
	$\beta$	12,1847	3,4689	<0,01 ***	[6,94; 26,42]
	$\gamma$	0,4501	0,0518	<0,01 ***	[0,35; 0,58]
<i>Von Bertalanffy</i>	$\alpha$	752,2838	30,3880	<0,01 ***	[698,31; 825,20]
	$\beta$	1,7729	0,3065	<0,01 ***	[1,21; 2,15]
	$\gamma$	0,3336	0,0365	<0,01 ***	[0,26; 0,39]
<i>Logístico</i>	$\alpha$	702,8716	13,9398	<0,01 ***	[673,98; 733,96]
	$\beta$	4,4426	0,3508	<0,01 ***	[3,78; 5,27]
	$\gamma$	0,6886	0,0574	<0,01 ***	[0,58; 0,82]

Fonte: Elaborada pela autora, (2024).

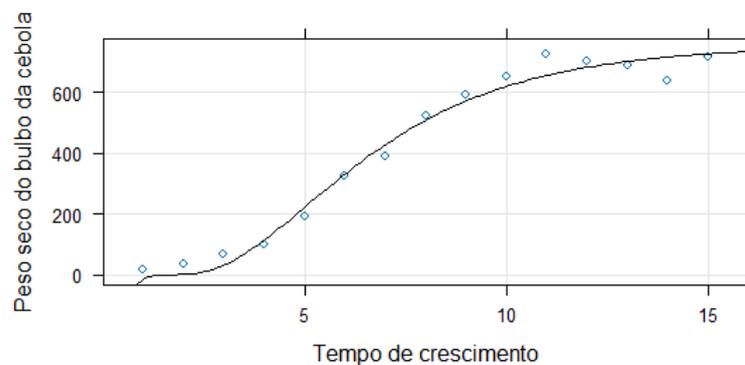
Observa-se, por meio da Tabela 4.2, que todos os parâmetros são estatisticamente significativos, ao nível de 5% adotado, para a variável resposta do modelo, conforme indicado pelos valores-p. Além disso, os intervalos de confiança apresentam-se bastante estreitos, o que sugere uma precisão elevada nas estimativas obtidas, indicando que os valores estimados refletem bem o efeito na população de origem. A seguir, nas Figuras 4.2-4.4, serão apresentados os gráficos de ajuste para os três modelos, ilustrando a relação entre o peso seco do bulbo da cebola e o tempo de crescimento.

Figura 4.2 – Gráfico de ajuste do modelo não linear Gompertz para o Peso seco do bulbo da cebola e Tempo de crescimento.



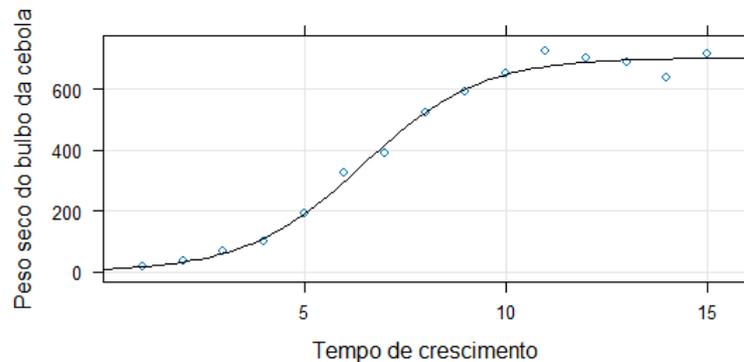
Fonte: Elaborada pela autora, (2024).

Figura 4.3 – Gráfico de ajuste do modelo não linear Von Bertalanffy para o Peso seco do bulbo da cebola e Tempo de crescimento.



Fonte: Elaborada pela autora, (2024).

Figura 4.4 – Gráfico de ajuste do modelo não linear Logístico para o Peso seco do bulbo da cebola e Tempo de crescimento.



Fonte: Elaborada pela autora, (2024).

Ao analisar os três gráficos, percebe-se que o modelo Logístico oferece o melhor ajuste para o banco de dados em questão. Para validar essa observação, foram aplicados os Critérios de Informação de Akaike (AIC) e Bayesiano de Schwarz (BIC). Na Tabela 4.3 são expostos os resultados detalhados dessas análises, confirmando a superioridade do modelo Logístico em termos de ajuste aos dados.

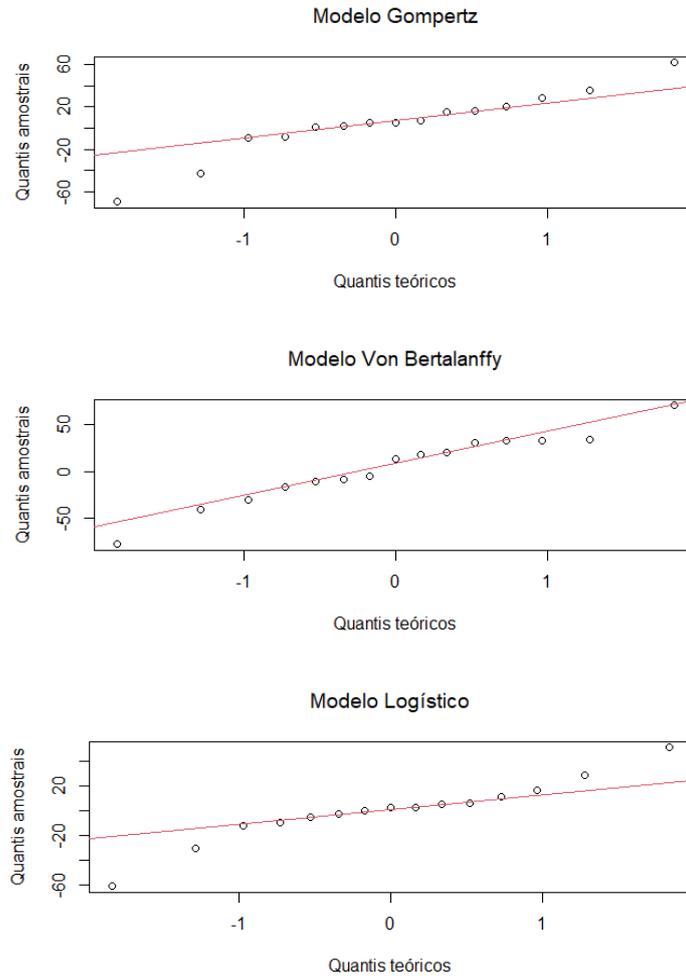
Tabela 4.3 – Critérios AIC e BIC para os modelos não lineares.

Modelo Não Linear	AIC	BIC
<i>Gompertz</i>	152,7216	155,5538
<i>Von Bertalanffy</i>	157,9225	160,7547
Logístico	146,4048	149,237

Fonte: Elaborada pela autora, (2024).

Pelos resultados constata-se claramente que, como os valores de AIC e BIC do modelo não linear Logístico são os menores, esse é o modelo mais apropriado para a análise do peso do bulbo em termos do tempo de crescimento.

Figura 4.5 – Normalidade dos Resíduos.



Fonte: Elaborada pela autora, (2024).

A Figura 4.5 apresenta os gráficos de normalidade dos resíduos para os modelos ajustados. Observa-se que, para o modelo Logístico, os pontos estão mais próximos da linha de referência, sugerindo uma distribuição aproximadamente normal dos resíduos. Esse comportamento indica que o modelo Logístico apresenta melhor aderência à suposição de normalidade dos resíduos, o que reforça sua adequação aos dados analisados. Em contrapartida, os modelos *Gompertz* e *Von Bertalanffy* apresentam desvios mais acentuados da linha de referência, o que pode sinalizar uma menor adequação desses modelos para capturar a dinâmica de crescimento do bulbo da cebola.

Tabela 4.4 – Teste de Shapiro-Wilk.

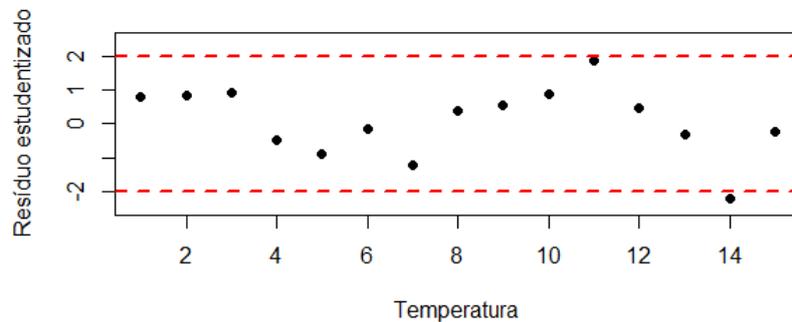
Modelo	p-valor
<i>Gompertz</i>	0,2558
<i>Von Bertalanffy</i>	0,7607
Logístico	0,2776

Fonte: Elaborada pela autora, (2024).

Os testes de Shapiro-Wilk mostram p-valores acima de 0.05 para os três modelos, indicando que os resíduos dos mesmos seguem uma distribuição aproximadamente normal. Isso significa que a condição de normalidade dos resíduos foi atendida, validando o uso desses modelos para analisar o crescimento do bulbo de cebola e permitindo comparações confiáveis entre eles quanto ao ajuste e à capacidade preditiva.

Na Figura 4.6 são expostos os resíduos estudentizados do modelo não linear Logístico ajustado para o crescimento do bulbo da cebola em função da temperatura.

Figura 4.6 – Gráfico dos resíduos estudentizados do modelo não linear Logístico.



Fonte: Elaborada pela autora, (2024).

Sabe-se que resíduos próximos à linha zero indicam que o modelo está ajustado adequadamente, o que é exibido na Figura 4.6. A distribuição aleatória dos pontos em torno da linha zero, sem formar padrões, sugere que o modelo está capturando bem a relação entre as variáveis. As linhas tracejadas vermelhas indicam limites de controle, e não há pontos fora desses limites, o que reforça a adequação do modelo.

## 4.2 Aprendizado de Máquina

No que diz respeito aos modelos gerados via *Machine Learning*, segue os resultados.

Tabela 4.5 – Comparação de medidas para os modelos de Aprendizado de Máquina.

Modelo de Aprendizado	RMSE	Acurácia	MAE
Árvore de Decisão	267,8882	0,0000	0,7509
<i>Random Forest</i>	0,2649	1,0000	0,2261
SVM	429,0254	1,0000	0,6393
XGBoost	0,3468	1,0000	0,2923

Fonte: Elaborada pela autora, (2024).

O modelo *Random Forest* se destacou como o melhor para prever o peso seco do bulbo da cebola, apresentando o melhor desempenho geral em todas as métricas analisadas. Com o menor valor de RMSE, o modelo demonstrou grande precisão nas previsões, refletindo uma diferença pequena entre os valores reais e previstos. Além disso, obteve 100% de acurácia na categorização dos dados, classificando corretamente os valores em baixo, médio e alto, e apresentou o menor valor de MAE (0,2261), o que indica que os erros médios nas previsões foram muito baixos. Em comparação, os outros modelos, como a Árvore de Decisão e o SVM, mostraram-se menos eficientes, com resultados significativamente mais elevados, tanto no RMSE quanto no MAE. O XGBoost também teve bom desempenho, mas ficou ligeiramente atrás do *Random Forest* em termos de precisão nas previsões. Portanto, o *Random Forest* foi o modelo mais eficaz para esse conjunto de dados, combinando alta acurácia, baixo erro médio e boa capacidade de previsão.

### 4.3 Comparação dos modelos

Tabela 4.6 – Comparação do RMSE e Acurácia entre Modelo Logístico e Random Forest.

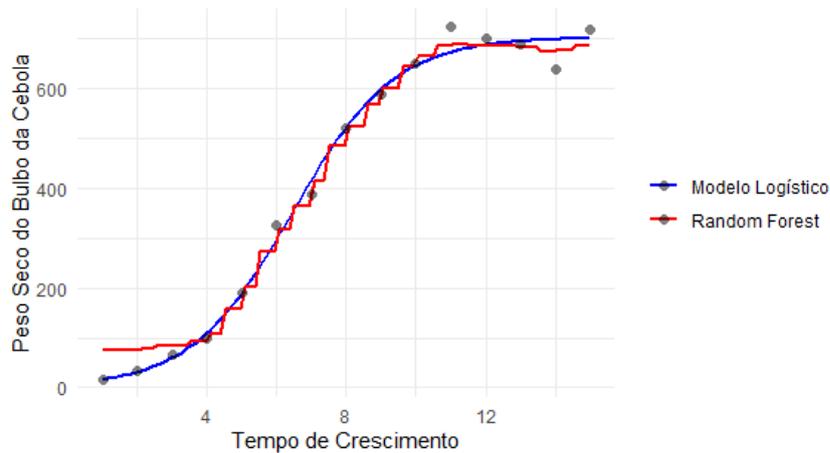
Modelo	RMSE	Acurácia
Logístico	24,3993	0,8666
<i>Random Forest</i>	0,2649	1,0000

Fonte: Elaborada pela autora, (2024).

O modelo Logístico apresenta um RMSE de 24,3993, indicando uma precisão relativamente baixa, com previsões distantes em média de 24,4 unidades do valor real. Embora esse erro seja inferior ao observado em outros modelos, como Árvore de Decisão e SVM, ele ainda sugere uma margem considerável de imprecisão. Em contrapartida, o modelo *Random Forest*, com um RMSE de apenas 0,2649, demonstra uma precisão significativamente superior, com previsões que estão a apenas 0,26 unidades do valor real. Isso indica que o *Random Forest* é muito mais eficaz em termos de acurácia nas previsões, sendo a escolha mais adequada para modelar o crescimento do bulbo de cebola, especialmente no que diz respeito à precisão.

Além disso, o modelo *Random Forest* se destaca ainda mais na avaliação de acurácia, atingindo 100%, enquanto o modelo Logístico obteve uma acurácia de 86,66%. Este desempenho superior do *Random Forest* evidencia sua capacidade de identificar de forma mais eficiente e precisa os padrões complexos e não lineares presentes nos dados de peso seco do bulbo da cebola e tempo de crescimento. Com base nesses resultados, é evidente que o *Random Forest* oferece uma solução mais robusta e confiável para modelar e prever o crescimento da cebola.

Figura 4.7 – Gráfico de comparação dos melhores modelos.



Fonte: Elaborada pela autora, (2024).

Observa-se, por meio da Figura 4.7, a precisão das previsões. O modelo *Random Forest* se ajusta de maneira mais próxima aos dados reais, refletindo seu menor RMSE. Embora o modelo Logístico seja eficiente, principalmente nos quatro primeiros dias de crescimento, o modelo *Random Forest* oferece uma predição ligeiramente mais precisa, especialmente após esse período, em que ocorre maior variação nos dados. Isso sugere que, em fases de crescimento mais avançadas, o *Random Forest* apresenta melhor desempenho para capturar as variações do peso seco do bulbo da cebola.

## 5 CONCLUSÃO

Neste trabalho, foram aplicados e comparados modelos de regressão não linear e algoritmos de Aprendizado de Máquina para descrever o crescimento do peso seco do bulbo da cebola (*Allium cepa L.*) em função do tempo de crescimento. Os resultados mostraram que os modelos não lineares avaliados, *Gompertz*, *Von Bertalanffy* e Logístico, apresentaram ajuste significativo, com destaque para o modelo Logístico, que se sobressaiu em termos de AIC e BIC. Esse modelo foi particularmente eficaz em descrever a curva de crescimento devido à sua capacidade de representar o comportamento biológico assintótico observado no cultivo da cebola.

A análise também demonstrou o potencial dos algoritmos de Aprendizado de Máquina, com o *Random Forest* destacando-se por sua superioridade preditiva. Este modelo apresentou o menor erro quadrático médio (RMSE) entre os métodos testados, mostrando-se mais adequado para capturar a variabilidade e a complexidade dos dados. A precisão das previsões obtidas pelo *Random Forest* sugere sua aplicabilidade prática em cenários agrícolas, onde condições heterogêneas e não linearidades são comuns.

Além de identificar os modelos mais adequados para descrever e prever o crescimento do bulbo da cebola, este estudo evidenciou a relevância de integrar abordagens tradicionais e modernas. Os modelos não lineares, embora sejam fundamentais para a interpretação biológica, podem ser complementados por algoritmos de Aprendizado de Máquina, capazes de lidar com grandes volumes de dados e interações complexas. Essa integração amplia o alcance das análises, permitindo previsões mais robustas e resultados valiosos para o manejo agrícola.

Por exemplo, as previsões obtidas podem auxiliar diretamente na tomada de decisões estratégicas, como o planejamento de colheitas, a determinação do momento ideal para intervenções no cultivo e a alocação eficiente de recursos. O uso do *Random Forest* para identificar padrões de crescimento pode ainda orientar práticas como a otimização da irrigação e o controle de fertilizantes, promovendo maior sustentabilidade e eficiência no cultivo.

Conclui-se que a abordagem adotada neste trabalho contribui tanto para a compreensão do crescimento da cebola quanto para o aprimoramento de práticas agrícolas. A integração de modelos não lineares, que fornecem interpretações biológicas robustas, com algoritmos de Aprendizado de Máquina, como o *Random Forest*, amplia as possibilidades de análise e previsão. Os resultados obtidos oferecem subsídios práticos para o manejo agrícola, como o planejamento de colheitas, a definição do momento ideal para intervenções e a otimização do uso de recursos, promovendo sustentabilidade e eficiência no cultivo da cebola.

## REFERÊNCIAS

- ALPAYDIN, E. **Introduction to Machine Learning**. [S.l.]: MIT Press, 2020. ISBN 978-0262039406.
- AWAD, N.; KHANNA, R. Support Vector Machines for Classification. In: **LEARNING from Imbalanced Data Sets**. Cham: Springer, 2015. P. 119–138. DOI: <10.1007/978-3-319-06766-4\_7>.
- BATES, Douglas M.; WATTS, Donald G. **Nonlinear Regression Analysis and Its Applications**. New York: Wiley, 1988.
- \_\_\_\_\_. Relative Curvature Measures of Nonlinearity. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 42, n. 1, p. 1–25, 1980.
- BEALE, E. M. L. Confidence Regions in Nonlinear Estimation. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 22, p. 41–88, 1960.
- BLEASDALE, J. K. A.; NELDER, J. A. Plant population and crop yield. **Nature**, v. 188, p. 342–347, 1960. DOI: <10.1038/188342a0>.
- BOX, George E. P. Bias in Nonlinear Estimation. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 33, p. 171–201, 1971.
- BOX, George E. P.; DRAPER, Norman R. **Response Surfaces, Mixture Designs, and Ridge Analyses**. New York: Wiley, 1987. ISBN 978-0471810339.
- BREIMAN, Leo. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. DOI: <10.1023/A:1010933404324>.
- BURNHAM, Kenneth P.; ANDERSON, David R. **Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach**. New York: Springer, 2002.
- CASELLA, George; BERGER, Roger L. **Statistical Inference**. 2nd. Pacific Grove, CA: Duxbury Press, 2002.
- CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A Scalable Tree Boosting System. In: **PROCEEDINGS of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. [S.l.: s.n.], 2016. P. 785–794. DOI: <10.1145/2939672.2939785>.
- COOK, R. Dennis; TSAI, T. Diagnostics for Mixed Models. **Journal of the American Statistical Association**, v. 80, n. 392, p. 258–263, 1985. DOI: <10.1080/01621459.1985.10478147>.
- COOK, R. Dennis; WEISBERG, Sanford. **Residuals and Influence in Regression**. Beverly Hills, CA: SAGE Publications, 1982.

- CORTES, Corinna; VAPNIK, Vladimir. Support-Vector Networks. **Machine Learning**, v. 20, n. 3, p. 273–297, 1995. DOI: {10.1007/BF00994018}.
- DANIEL, Fabrício. A Survey on Machine Learning Applications. **Journal of Artificial Intelligence Research**, v. 71, p. 1–34, 2021.
- DEMING, W. Edwards. **Out of the Crisis**. Cambridge, MA: MIT Press, 1986. A obra aborda princípios de melhoria contínua e gestão de qualidade.
- DING, Wei; CAO, Hu; NÆSS, Lars. Gradient Boosting Machine: An Overview. **Journal of Machine Learning Research**, v. 22, n. 61, p. 1–20, 2021.
- EVANS, G. C. **The quantitative analysis of plant growth**. London: Blackwell Scientific Publications, 1972.
- FARIA, Marcelo; MAY, A.; RESENDE, Julio Tadeu V.; GALVÃO, Anízio Gonçalves; MORAES, A. R. A.; RIBEIRO, W. S. Produtividade e rendimento comercial de bulbos de cebola em função da densidade de cultivo. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 10, n. 4, p. 959–963, 2006. DOI: {10.1590/S1415-43662006000400021}.
- FLETCHER, Roger. **Practical Methods of Optimization**. 2nd. Chichester, UK: Wiley, 1987.
- GALLANT, A. Ronald. **Nonlinear Statistical Models**. New York: Wiley, 1987.
- GILL, Philip E.; MURRAY, Walter; WRIGHT, Margaret H. **Practical Optimization**. London: Academic Press, 1981.
- GILLIS, James; RATKOWSKY, David A. The Behavior of Estimators of Nonlinear Models in the Presence of Multicollinearity. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 40, p. 362–374, 1978.
- GUIMARÃES, R. J. **Análise do crescimento e da quantificação de nutrientes em mudas de cafeeiro (Coffea arabica L.), durante seus estádios de desenvolvimento em substrato padrão**. 1994. Dissertação de Mestrado – Escola Superior de Agricultura de Lavras (ESAL), Lavras, MG. Mestrado em Fitotecnia.
- HOLLIDAY, R.; ADAMS, M. W.; ANDERSON, H. R. The growth of the tobacco plant and its relation to nitrogen supply. **Annals of Botany**, v. 24, p. 147–169, 1960.
- MATTOS, Sergio L. **Análise de Regressão: Teoria e Aplicações**. São Paulo: Editora Atlas, 2013. P. 29.
- MONTGOMERY, Douglas C.; PECK, Elizabeth A. **Introduction to Linear Regression Analysis**. 2nd. New York: Wiley, 1992. ISBN 978-0471850656.
- MYERS, Raymond H. **Classical and Modern Regression with Applications**. 2nd. Boston: Duxbury Press, 1990. ISBN 978-0534929858.

MYERS, Robert H. *Classical and Modern Regression with Applications*. Duxbury Press, 1990. Richards' growth model discussion.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). **NIST Data Repository**. [S.l.: s.n.], 2024. Acessado em: 09 de novembro de 2024. Disponível em: [i{https://www.nist.gov/}](https://www.nist.gov/)i.

NILSSON, Nils J. **Artificial Intelligence: A New Synthesis**. San Francisco: Morgan Kaufmann, 1998.

NOCEDAL, Jorge; WRIGHT, Stephen J. **Numerical Optimization**. 2nd. New York: Springer, 2006.

POMEROY, Laurence; WILLIAMS, John; MOORE, Sarah. The Logistic Growth Model: Applications and Extensions in Population Dynamics. **Ecological Modelling**, v. 101, p. 345–358, 1997.

QUINLAN, J. R. Induction of Decision Trees. **Machine Learning**, v. 1, n. 1, p. 81–106, 1986. DOI: [10.1023/A:1022643204877](https://doi.org/10.1023/A:1022643204877).

RATKOWSKY, David A. **Nonlinear Regression Modeling: A Unified Practical Approach**. New York: Marcel Dekker, 1983.

RICKER, W. E. Stock and recruitment. **Journal of the Fisheries Research Board of Canada**, v. 11, p. 559–623, 1954. DOI: [10.1139/f54-039](https://doi.org/10.1139/f54-039).

RYO, Masashi; RILLIG, Matthias C. Machine Learning Approaches to Predicting Plant Growth. **Trends in Plant Science**, v. 22, n. 11, p. 893–905, 2017. DOI: [10.1016/j.tplants.2017.06.008](https://doi.org/10.1016/j.tplants.2017.06.008).

SAMOLETOV, Andrey; IVANOV, Sergey; POPOV, Maxim. A Study on the Gompertz Growth Model for Predicting Population Dynamics. **International Journal of Mathematical Models and Methods in Applied Sciences**, v. 11, p. 123–130, 2017.

SEBER, George A. F.; WILD, C. J. **Nonlinear Regression**. Hoboken, NJ: Wiley-Interscience, 2003.

SHALEV-SHWARTZ, Shai; BEN-DAVID, Shai. **Understanding Machine Learning: From Theory to Algorithms**. Cambridge: Cambridge University Press, 2011.

TEAM, R Core. **R: A Language and Environment for Statistical Computing**. [S.l.: s.n.], 2024. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: [i{https://www.R-project.org/}](https://www.R-project.org/)i.

UEDA, Takashi. **Model Selection and Model Averaging**. Tokyo: Springer, 2003. P. 55.

VON BERTALANFFY, Ludwig. **Untersuchungen über die Gesetzlichkeit des Wachstums. II. Wachstumsgesetze**. Berlin: De Gruyter, 1934.