

Lucas da Mata Guimarães

## **Titulo do Trabalho**

São Paulo - Brasil

2025

Lucas da Mata Guimarães

## **Titulo do Trabalho**

Monografia apresentada na disciplina Trabalho de Conclusão de Curso, como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação.

Centro Universitário Senac - Santo Amaro  
Bacharelado em Ciência da Computação

Orientador: Nome do Orientador

São Paulo - Brasil  
2025

*Texto da dedicatória.*

# Agradecimentos

Texto de agradecimento.

*“A vingança nunca é plena,  
mata a alma e a envenena.  
(MADRUGA, Seu, Chaves)*

# Resumo

Texto do resumo

**Palavras-chaves:** palavra-chave 1, palavra-chave 2, palavra-chave 3.

# Abstract

Abstract text in english

**Key-words:** keyword 1, keyword 2, keyword 3

# Lista de ilustrações

Figura 1 – Regressão Linear Simples . . . . .	16
Figura 2 – Regressão Linear Multipla . . . . .	18



# Lista de tabelas

Tabela 1 – Matriz de Confusão . . . . .	18
---	----

# Lista de abreviaturas e siglas

GAM	Generalized Additive Models
GLM	Generalized Linear Model
MARS	Multivariate Adaptive Regression Spline
ML	Maximum likelihood

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>11</b>
<b>1.1</b>	<b>Contexto</b>	<b>11</b>
<b>1.2</b>	<b>Justificativa</b>	<b>11</b>
<b>1.3</b>	<b>Objetivo</b>	<b>12</b>
1.3.1	Objetivos Específicos	12
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>14</b>
<b>2.1</b>	<b>Modelos Computacionais</b>	<b>14</b>
2.1.1	Modelos Lineares	14
2.1.2	Acurácia	17
2.1.3	Modelos de Distribuição de Espécies	19
2.1.3.1	GLM - Generalized Linear Model	20
2.1.3.2	GAM - Generalized Additive Model	20
2.1.3.3	MARS - Multivariate Adaptive Regression Spline	20
2.1.4	Análise Regressiva	20
2.1.5	Maximum Likelihood	20
<b>2.2</b>	<b>Análise de Algoritmos</b>	<b>20</b>
2.2.1	Análise de Complexidade	20
2.2.2	Análise de Espaço	20
<b>2.3</b>	<b>Análise de Dados em larga escala</b>	<b>20</b>
<b>2.4</b>	<b>Linguagem R</b>	<b>20</b>
2.4.1	Bibliotecas	20
<b>2.5</b>	<b>Trabalhos relacionados</b>	<b>20</b>
<b>3</b>	<b>DESENVOLVIMENTO</b>	<b>21</b>
<b>4</b>	<b>RESULTADOS</b>	<b>22</b>
<b>5</b>	<b>CONCLUSÃO</b>	<b>23</b>
<b>5.1</b>	<b>Trabalhos Futuros</b>	<b>23</b>
	<b>REFERÊNCIAS</b>	<b>24</b>
	<b>APÊNDICES</b>	<b>27</b>
	<b>APÊNDICE A – EXEMPLO DE SEÇÃO DE ANEXO</b>	<b>28</b>

# 1 Introdução

## 1.1 Contexto

O uso de modelos computacionais, na Biologia, possibilita o avanço de diferentes estudos ([COSME, 2025a](#)). Uma destas aplicações são os modelos de distribuição de espécies, que são capazes de fornecer uma visualização da situação da fauna e flora de determinada região, podendo mostrar como estas estão se comportando no decorrer do tempo ([ELITH; LEATHWICK, 2009](#)).

Entre esses modelos, os mais utilizados são o Generalized Additive Models (GAM) ([HASTIE; TIBSHIRANI, 1986](#)) e o Generalized Linear Model (GLM) ([PAUL; SAHA, 2007](#)). Esses dois modelos usam uma função para estabelecer uma relação entre a média da variável de resposta e uma função 'suavizada' das variáveis explanatórias, sendo o GLM uma extensão de modelos lineares que não forçam o dado a escalas não naturais, e o GAM uma extensão semi-parametrizada do GLM, tendo a capacidade de atuar com relações não lineares e não monótonas ([GUISAN; EDWARDS; HASTIE, 2002](#)).

Já o Multivariate Adaptive Regression Spline (MARS) combina partição recursiva e ajustes por splines, de modo a manter seus aspectos positivos, enquanto sendo menos vulnerável a suas propriedades não favoráveis. Gerando um conjunto de regras para prever valores futuros apartir de uma análise regressiva. ([FRIEDMAN, 1991](#))

Sendo as aplicações destes modelos encontradas codificadas na linguagem de programação R, que por sua vez é a linguagem de programação mais utilizada quando tratamos de ciência de dados, sendo conhecida como a linguagem mais robusta para a área de dados, tendo sido pensada para o uso em cálculos e análises estatísticas ([AWARI, 2022](#)).

Porém, estes modelos podem requisitar uma alta demanda de processamento e memória do computador hospedeiro, como citado por ([COSME, 2025a](#)), ponto este, que não é repassado nos trabalhos referentes a análise ou uso dos modelos citados. Logo, mesmo com a facilidade de se adquirir um computador, tais modelos requerem computadores de alto desempenho para serem treinados, tornando esse processo lento ou criando a necessidade de se alugar máquinas virtuais para esta finalidade ([RICHTER, 2025](#)).

E quando se coloca a necessidade de se manter um controle das populações de espécies, dentro ou próximo a centros urbanos, a velocidade de preparo destes modelos se torna mais crítica, já que é necessário ir desde a coleta dos dados, ao treino e validação do modelo, e análise dos resultados obtidos.

## 1.2 Justificativa

Identificar a distribuição de espécies em um dado ambiente, em um determinado intervalo de tempo, é importante para termos noção de como as espécies estão respondendo a mudanças no ambiente, no aumento ou diminuição de outra espécie.

Uma vez que essas mudanças podem ser geradas pela ação humana, na construção civil e de infraestrutura ([AMETEPEY; ANSAH, 2014](#)), conseguir estimar o impacto dessas

ações é vantajoso para a preservação de espécies.

Além disso, estas abordagens aumentam as possibilidades para integrar a infraestrutura necessária, contribuindo para a sobrevivência de espécies que estão em níveis populacionais baixos.

Modelos estatísticos, que tem a capacidade de demonstrar estes eventos, aplicam de maneiras diferentes algumas linhas de abordagem. O Generalized Additive Models (GAM), Generalized Linear Model (GLM), e o Multivariate Adaptive Regression Spline (MARS), ambos com uma abordagem de Maximum likelihood (ML), variando em sua capacidade de atuar com um determinado tipo de dado e o custo levado para seu treinamento e utilização (NORBERG et al., 2019).

Modelos que são utilizados na modelagem de distribuição de espécies necessitam de uma quantidade elevada de dados (WISZ et al., 2008), de ocorrência e ausência, sendo os dados de ausência não necessários em todos os tipos de modelos.

Nem todas as espécies são facilmente modeláveis devido à dificuldade de coleta de dados, seja pela sua raridade ou habitat (STOCKMAN; BEAMER; BOND, 2006). A colaboração de cidadãos na coleta de dados pode auxiliar na identificação de áreas prioritárias para pesquisa. Portanto, a identificação de bons modelos que trabalham com esses dados é vantajosa.

Dentro destes modelos, além da quantidade e tipo de dados necessários, precisamos levar em consideração, o custo necessário de processamento e o espaço de memória utilizado pelo mesmo, para este fim utilizamos a análise de complexidade e espaço (CORMEN et al., 2009), já que um modelo mais barato nesse quesito pode ser criado em computadores mais acessíveis (SEDGEWICK; FLAJOLET, 2013), e ser possível a construção de mais de um modelo de modo simultâneo.

Os pontos levantados anteriormente podem afetar a acurácia de um modelo, mesmo atendendo os requisitos, de pouco adianta se o mesmo nos entrega respostas que induzem ao erro. Identificar um modelo que tenham uma boa acurácia, quando trabalham somente com dados de ocorrência, assim como uma melhor avaliação computacional, se vê vantajoso para situações em que queremos criar uma análise inicial de um determinado cenário.

## 1.3 Objetivo

Este trabalho tem como objetivo avaliar e comparar a implementação encontrada nas bibliotecas mda e mgcv da linguagem R, dos modelos de distribuição de espécies, GAM, GLM e MARS, levantando o custo computacional de cada um destes apartir de uma análise de complexidade e espaço. Encontrando um modelo que melhor apresente um equilíbrio entre a acurácia e o custo computacional.

### 1.3.1 Objetivos Específicos

1. Análise de complexidade e espaço dos modelos.

- Generalized Additive Model;
- Generalized Linear Model;
- Multivariate Adaptive Regression Spline;

2. Avaliação da acurácia dos modelos com dados de ocorrência.
3. Comparação dos modelos.
4. Avaliação dos modelos com base na relação custo x acurácia.

## 2 Revisão Bibliográfica

### 2.1 Modelos Computacionais

Modelos computacionais são modelos que representam fenômenos de modo simplificado, gerando uma aproximação do evento real, tendo em vista a visualização ou entendimento de determinado fenômeno, codificados em alguma linguagem computacional para ser executado em um computador. Estes modelos podem ser criados por especialistas utilizando de equações matemáticas ou, automaticamente utilizando de técnicas de inteligência artificial. (AUGUSTO, 2025)

Ao processo de criação destes modelos, damos o nome de modelagem computacional, podendo ser aplicado em qualquer situação onde uma análise de um sistema complexo se vê necessária, sendo suas principais aplicações encontradas nas seguintes áreas, como apresentado por (COSME, 2025b):

1. **Ciência e Pesquisa:** Permite o teste de hipóteses de maneira mais rápida e eficiente.
2. **Engenharia:** Essencial para projetos de larga escala, utilizada para testar estruturas antes de começar sua construção.
3. **Medicina:** Permite a modelagem de epidemias, assim prevendo como doenças podem se espalhar em dada população, ajudando a planejar métodos de controle.

O tipo da modelagem depende do tipo de fenômeno ou problema que queremos tratar, onde os tipos principais segundo (COSME, 2025b) são:

1. **Modelagem determinística:** O comportamento do sistema é previsível, onde os mesmos parâmetros de entrada sempre produzem os mesmos resultados. Mais visto no campo da Física e Engenharia, onde os fenômenos naturais seguem um conjunto de regras bem definido.
2. **Modelagem estocástica:** Inclui elementos de incerteza e aleatoriedade, o sistema pode apresentar resultados diferentes para o mesmo conjunto de parâmetros de entrada. Comumente usada onde o acaso desempenha um papel importante, como na Biologia e Economia.
3. **Modelagem dinâmica:** Focada em sistemas que mudam ao longo do tempo, essencial em áreas como a Ecologia e Epidemiologia, onde se é preciso prever a evolução de sistemas biológicos ou a propagação de doenças.

#### 2.1.1 Modelos Lineares

Modelos lineares são modelos que preveem uma resposta linear utilizando de base a relação entre o resultado e as propriedades dadas como parâmetros. Sendo uma opção

mais simples, possui propriedades mais fáceis de serem entendidas e um tempo de desenvolvimento mais curto quando comparado a outros tipos de modelos, como redes neurais, ou árvores de decisão, empregadas no mesmo problema. (IBM, 2025)

A linearidade destes modelos, implica que matematicamente a variação dos parâmetros independentes não possuem relações entre si, e podem ser separados em dois grupos clássicos (ADALARDO, 2020).

- **Modelos de Regressão:** Este grupo é utilizado para modelar relações entre variáveis quantitativas, que são um conjunto de valores de possível representação numérica, indicando quantidade ou magnitude. Com o intuito de estimar parâmetros, explicando relação ou para fazer previsões.
- **Modelos de Análise de Variância:** Estes modelos tem como questão principal comparar a importância de fatores sobre o comportamento da variável de resposta. Para encontrar a relação entre grupos de análise, de modo a indentificar oque gera a diferença entre os grupos estudados.

Ambas as abordagens ao modelo linear, irão gerar uma regressão linear, que é um modelo matemático que descrevem a relação entre as váriaveis dependentes e independentes usadas, tendo a possibilidade de ser representado graficamente. Podendo ser de dois tipos, simples ou múltipla.

Na regressão linear simples, queremos estimar os valor de  $a$  e  $b$  da equação da reta,  $y = a + bx$ , apartir de um conjunto de dados  $x$  e  $y$ , onde  $y$  representa a váriavel dependete e  $x$  á váriavel independente, que melhor represente a relação entre  $x$  e  $y$ . Em outras palavras, queremos estimar a inclinação da reta, esta que nos indica o efeito em  $y$  das mudanças ocorridas em  $x$  (CHEIN, 2019).

A essa reta, é dado o nome de reta de regressão linear, está que depende de cinco estatísticas básicas (CHEIN, 2019):

1. Média de  $X$ :  $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ ;
2. Desvio padrão de  $X$ :  $S_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$ ;
3. Média de  $Y$ :  $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$ ;
4. Desvio padrão de  $Y$ :  $S_y = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2}$ ;
5. Correlação de  $X$  e  $Y$ :  $r = \frac{1}{n} \sum_{i=1}^N \frac{X_i - \bar{X}}{S_x} \cdot \frac{Y_i - \bar{Y}}{S_y}$

Com estas estatísticas podemos traçar a reta de regressão, sabendo que esta passa pelo ponto médio  $(\bar{X}, \bar{Y})$ . A inclinação da reta será dada por:

$$\beta_1 = \frac{r \cdot S_y}{S_x} \quad (2.1)$$

E o intercepto da reta de regressão, onde a reta corta um dos eixos do plano cartesiano, será dado por:



$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \quad (2.2)$$

Assim resultamos na seguinte equação:

$$Y = \beta_0 + \beta_1 X \quad (2.3)$$

Onde:

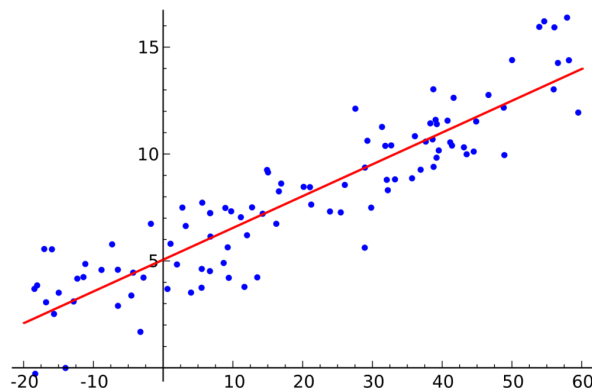
- $(X)$  é a variável independente;
- $(Y)$  é a variável dependente;
- $(\beta_0)$  é o intercepto da reta;
- $(\beta_1)$  é a inclinação da reta.

Porém, a equação 2.3 ainda não proporciona os valores de  $Y$ , mesmo possuindo os valores para  $\beta_0$  e  $\beta_1$ , visto que não é apenas a variável  $X$  que afeta os valores de  $Y$  quando tratamos de ocorrência no mundo real, assim incluímos um termo de erro  $\epsilon$ , que é o erro que se comete ao estimar os valores de  $Y$  por meio da variável  $X$  (CHEIN, 2019).

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (2.4)$$

Agora com a equação 2.4, podemos criar a reta de regressão, que pode ser representada graficamente, possuindo uma estrutura semelhante ao gráfico a seguir:

Figura 1 – Regressão Linear Simples



Fonte: EBAC (2023)

A equação 2.4, pode ser escrita de forma mais geral. Visto que em nossos dados podemos trabalhar com conjuntos de valores, que agrupam valores de  $Y$  distintos, para cada valor de  $X$ . Por exemplo, dados que representam a qualidade de vida nos estados brasileiros com o número de postos de saúde. Assim a equação 2.4 ficaria (CHEIN, 2019):

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (2.5)$$

Onde:

- $(i)$  representa um agrupamento de dados, um estado brasileiro seguindo o exemplo dado a sima;
- $(Y_i)$  é a variável dependente;
- $(X_i)$  é a variável independente, representaria o número de postos de saúde em dado estado.
- $(\beta_0)$  é o intercepto;
- $(\beta_1)$  é a inclinação, e o efeito médio de  $X_i$  sobre  $Y_i$ ;
- $(\epsilon)$  é o erro médio ao se estimar  $Y_i$  por meio de  $X_i$ ;

Já quando tratamos da regressão linear múltipla, é levado em conta que outros fatores podem afetar a variável de resposta, este que podem também ser correlacionados com a variável independente. A formula para este modelo de regressão, pode ser representada da seguinte forma, onde temos  $k$  variáveis, explicativas (CHEIN, 2019):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon \quad (2.6)$$

Onde:

- $(\beta_0)$  é o intercepto;
- $(\beta_1, \dots, \beta_k)$  são "inclinações", mesmo que na prática não sejam inclinações da função
- $(\epsilon)$  é o termo de erro.

Aqui temos  $\beta_1$  até  $\beta_k$  como coeficientes parciais da regressão (CHEIN, 2019). Neste caso, a visualização por meio de um gráfico, fica comprometida, visto que temos um número  $k$  de  $X$ , para ilustrar, usemos uma situação onde temos dois  $X$ , aqui podemos representar os valores por um gráfico de três dimensões.

Quando temos mais de dois valores de  $X$  a representação gráfica fica confusa para o entendimento humano, mas ainda podemos tratar o problema como uma reta.

### 2.1.2 Acurácia

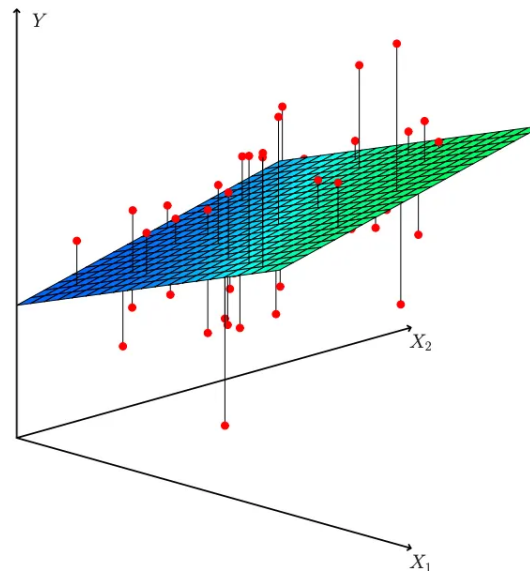
A acurácia é um modo de avaliar a performance de um modelo, assim identificando se seus resultados podem ser considerados válidos ou não. Para chegarmos na acurácia utilizamos da seguinte formula (FILHO, 2023):

$$A = \frac{PC}{TP} \quad (2.7)$$

Onde:

- $(A)$  é a Acurácia
- $(PC)$  é o total de Previsões Corretas, encontrada pela soma,  $(PositivosVerdadeiros + NegativosVerdadeiros)$ ;

Figura 2 – Regressão Linear Multipla



Fonte: [Novak \(2022\)](#)

- $(TP)$  é o Total de Previsões, encontrada pela soma,  $(PC + FalsosVerdadeiros + FalsosNegativos)$ .

Como a acurácia incorpora por completo a matriz de confusão [1](#), em um conjunto de dados equilibrado, com uma quantidade de exemplos semelhante para as duas classes, ela pode ser usada como uma métrica grosseira da qualidade de um modelo ([DEVELOPERS, 2024](#)).

Onde temos mais exemplos de uma classe do que de outras, é importante considerar outras métricas de avaliação, já que esses modelos são considerados desbalanceados ([FILHO, 2023](#)).

Uma matriz de confusão, por sua vez, é uma representação binária em forma de matriz dos resultados possíveis ao se utilizar o modelo, sendo representada da seguinte forma:

Tabela 1 – Matriz de Confusão

	Verdadeiro Positivo	Verdadeiro Negativo
Positivo Previsto	Positivo Verdadeiro	Falsos Positivos
Negativo Previsto	Falsos Negativos	Negativos Verdadeiros

Onde:

- **Positivos Verdadeiros:** São as classificações positivas corretas;
- **Falsos Positivos:** São positivos que foram erroneamente classificados como negativos;
- **Falsos Negativos:** São negativos que foram erroneamente classificados como positivos;
- **Negativos Verdadeiros:** São as classificações negativas corretas;

### 2.1.3 Modelos de Distribuição de Espécies

Definimos um Modelos de Distribuição de Espécies, SDM (Spice Distribution Model), como um modelo que relaciona dados de distribuição de espécies, com informações sobre as características ambientais e/ou epaciais de certas localidades, podendo ser usados para entender e/ou prever a distribuição de uma espécie por uma dada localidade ([ELITH; LEATHWICK, 2009](#)).

SDMs contemporaneos combinan conceitos de ecologia e historia natural com os avanços mais recentes em estatísticas e tecnologia da informação, as raízes destes modelos são encontradas no estudos mais antigos que descrevem padrões biológicos em termos de relações com geografia e/ou gradientes ambientais, e estudos que indicam a resposta individual de espécies para seus ambientais prove um forte argumento conceitual para se modelar espécies de modo individual ([ELITH; LEATHWICK, 2009](#)).

As principais fontes de inforamções para a criação destes modelos, são ([OLIVER, 2024](#)):

- Dado de ocorrência: Geralmente coordenadas de latitude e longitude onde a especie foi observada, conhecida como dado de **ocorrência**, alguns modelos faze uso de dados de **ausência**, que são coordenadas geograficas onde se sabe que a espécie não ocorre;
- Dado ambiental: São a descrição do ambiente, podendo conter medições de temperatura e precepitação, como também, a ocorrência e ausência de outras espécies, como predadores, competidores ou fontes de alimento.

Dentro dos SDMs, temos varios frameworks de modelagem, sendo os mais utilizados o Generalized Linear Model (GLM), Generalized Addtive Model (GAM) e Multivariate Adaptive Regression Spline (MARS), que são encontradas nos softwares mais amigaveis ao usuário e bem documentados ([NORBERG et al., 2019](#)), assim como podem ser encontrados em bibliotecas de linguagens de programação, como R nas bibliotecas mda, onde encontramos o GLM e GAM ([HASTIE et al., 2024](#)) e mgcv onde encontramos o MARS([WOOD, 2025](#)).

2.1.3.1 GLM - Generalized Linear Model

2.1.3.2 GAM - Generalized Additive Model

2.1.3.3 MARS - Multivariate Adaptive Regression Spline

2.1.4 Análise Regressiva

2.1.5 Maximum Likelihood

2.2 Análise de Algoritmos

2.2.1 Análise de Complexidade

2.2.2 Análise de Espaço

2.3 Análise de Dados em larga escala

2.4 Linguagem R

2.4.1 Bibliotecas

2.5 Trabalhos relacionados

### 3 Desenvolvimento

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed sollicitudin tempor sapien in maximus. Quisque in vulputate dui, ac vestibulum sem. Suspendisse urna velit, dapibus nec egestas a, rhoncus vitae neque. Mauris quis efficitur augue. Aliquam quis tellus eget orci aliquet aliquam. Sed luctus, quam vitae elementum malesuada, quam lacus imperdiet urna, sed ullamcorper libero magna non elit. Cras laoreet arcu a augue volutpat, suscipit pretium tellus tempus. Sed eros tortor, imperdiet eu neque id, interdum egestas tortor.

## 4 Resultados

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed sollicitudin tempor sapien in maximus. Quisque in vulputate dui, ac vestibulum sem. Suspendisse urna velit, dapibus nec egestas a, rhoncus vitae neque. Mauris quis efficitur augue. Aliquam quis tellus eget orci aliquet aliquam. Sed luctus, quam vitae elementum malesuada, quam lacus imperdiet urna, sed ullamcorper libero magna non elit. Cras laoreet arcu a augue volutpat, suscipit pretium tellus tempus. Sed eros tortor, imperdiet eu neque id, interdum egestas tortor.

## 5 Conclusão

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed sollicitudin tempor sapien in maximus. Quisque in vulputate dui, ac vestibulum sem. Suspendisse urna velit, dapibus nec egestas a, rhoncus vitae neque. Mauris quis efficitur augue. Aliquam quis tellus eget orci aliquet aliquam. Sed luctus, quam vitae elementum malesuada, quam lacus imperdiet urna, sed ullamcorper libero magna non elit. Cras laoreet arcu a augue volutpat, suscipit pretium tellus tempus. Sed eros tortor, imperdiet eu neque id, interdum egestas tortor.

### 5.1 Trabalhos Futuros

- Trabalho Futuro 1
- Trabalho Futuro 2
- Trabalho Futuro 3



# Referências

- ADALARDO. 7. *Modelos Lineares*. 2020. Acesso em: 26 de Abril de 2025. Disponível em: <[http://ecor.ib.usp.br/doku.php?id=03\\_apostila:06-modelos#:~:text=SÃo%20chamados%20modelos%20lineares%20aqueles,dos%20demais%20parÃmetros%20do%20modelo.](http://ecor.ib.usp.br/doku.php?id=03_apostila:06-modelos#:~:text=SÃo%20chamados%20modelos%20lineares%20aqueles,dos%20demais%20parÃmetros%20do%20modelo.)> Citado na página 15.
- AMETEPEY, S. O.; ANSAH, S. K. Impacts of construction activities on the environment : the case of ghana. *Journal of Construction Project Management and Innovation*, v. 4, n. sup-1, p. 934–948, 2014. Disponível em: <<https://journals.co.za/doi/abs/10.10520/EJC162729>>. Citado na página 11.
- AUGUSTO, D. A. *Entenda o que são modelos computacionais e como o SISS-Geo os utiliza*. 2025. Acesso em: 25 de Abril de 2025. Disponível em: <<https://www.biodiversidade.ciss.fiocruz.br/entenda-o-que-sao-modelos-computacionais-e-como-o-siss-geo-os-utiliza>>. Citado na página 14.
- AWARI. *Conheça as principais linguagens de programação para Ciência de Dados*. 2022. Acesso em: 25 de Abril de 2025. Disponível em: <<https://awari.com.br/linguagens-de-programacao-para-ciencia-de-dados/>>. Citado na página 11.
- CHEIN, F. *Introdução aos Modelos de Regressão Linear*. Enap, 2019. ISBN 9788525601155. Disponível em: <[https://repositorio.enap.gov.br/bitstream/1/4788/1/Livro\\_RegressÃo%20Linear.pdf](https://repositorio.enap.gov.br/bitstream/1/4788/1/Livro_RegressÃo%20Linear.pdf)>. Citado 3 vezes nas páginas 15, 16 e 17.
- CORMEN, T. et al. *Introduction to Algorithms, third edition*. MIT Press, 2009. (Computer science). ISBN 9780262033848. Disponível em: <<https://books.google.com.br/books?id=i-bUBQAAQBAJ>>. Citado na página 12.
- COSME, A. L. *Modelagem computacional: o que é, qual sua aplicação*. 2025. Acesso em: 17 de Abril de 2025. Disponível em: <<https://123ecos.com.br/docs/modelagem-computacional/>>. Citado na página 11.
- COSME, A. L. *Modelagem computacional: o que é, qual sua aplicação*. 2025. Acesso em: 25 de Abril de 2025. Disponível em: <<https://123ecos.com.br/docs/modelagem-computacional/#:~:text=Os%20principais%20tipos%20sÃo%20a,em%20diferentes%20Ãreas%20do%20conhecimento.>> Citado na página 14.
- DEVELOPERS, G. for. *Classificação: precisão, recall, precisão e métricas relacionadas*. 2024. Acesso em: 3 de Maio de 2025. Disponível em: <<https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall?hl=pt-br>>. Citado na página 18.
- EBAC. *Regressão Linear: teoria e exemplos*. 2023. Acesso em: 27/04/2025. Disponível em: <<https://ebaconline.com.br/blog/regressao-linear-seo>>. Citado na página 16.
- ELITH, J.; LEATHWICK, J. R. Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, Annual Reviews, v. 40, n. Volume 40, 2009, p. 677–697, 2009. ISSN 1545-2069. Disponível

em: <<https://www.annualreviews.org/content/journals/10.1146/annurev.ecolsys.110308.120159>>. Citado 2 vezes nas páginas 11 e 19.

FILHO, M. *O Que É Acurácia Em Machine Learning?* 2023. Acesso em: 3 de Maio de 2025. Disponível em: <<https://mariofilho.com/o-que-e-acuracia-em-machine-learning/>>. Citado 2 vezes nas páginas 17 e 18.

FRIEDMAN, J. H. Multivariate Adaptive Regression Splines. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 19, n. 1, p. 1 – 67, 1991. Disponível em: <<https://doi.org/10.1214/aos/1176347963>>. Citado na página 11.

GUISAN, A.; EDWARDS, T. C.; HASTIE, T. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, v. 157, n. 2, p. 89–100, 2002. ISSN 0304-3800. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304380002002041>>. Citado na página 11.

HASTIE, T.; TIBSHIRANI, R. Generalized Additive Models. *Statistical Science*, Institute of Mathematical Statistics, v. 1, n. 3, p. 297 – 310, 1986. Disponível em: <<https://doi.org/10.1214/ss/1177013604>>. Citado na página 11.

HASTIE, T. et al. *Package 'mda'*. 0.5-5. ed. [S.l.], 2024. <https://CRAN.R-project.org/package=mda>. Disponível em: <<https://cran.r-project.org/web/packages/mda/mda.pdf>>. Citado na página 19.

IBM. *Modelos lineares*. 2025. Acesso em: 26 de Abril de 2025. Disponível em: <<https://www.ibm.com/docs/pt-br/spss-modeler/18.5.0?topic=node-linear-models>>. Citado na página 15.

NORBERG, A. et al. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, v. 89, n. 3, p. e01370, 2019. Disponível em: <<https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecm.1370>>. Citado 2 vezes nas páginas 12 e 19.

NOVAK, H. *Modelos de Regressão Linear e como aplicamos a técnica na Loft*. 2022. Acesso em: 27/04/2025. Disponível em: <<https://medium.com/loftbr/regressão-linear-65fc8caeb729>>. Citado na página 18.

OLIVER, J. 2024. Acesso em: 4 de Maio de 2025. Disponível em: <[https://jcoliver.github.io/learn-r/011-species-distribution-models.html#:~:text=\(%20predicts%20-,Components%20of%20the%20model,these%20data%20\(see%20below\).>](https://jcoliver.github.io/learn-r/011-species-distribution-models.html#:~:text=(%20predicts%20-,Components%20of%20the%20model,these%20data%20(see%20below).>)>. Citado na página 19.

PAUL, S.; SAHA, K. K. The generalized linear model and extensions: a review and some biological and environmental applications. *Environmetrics*, v. 18, n. 4, p. 421–443, 2007. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/env.849>>. Citado na página 11.

RICHTER, F. *Amazon and Microsoft Stay Ahead in Global Cloud Market*. 2025. Acesso em: 19 de Abril de 2025. Disponível em: <<https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/>>. Citado na página 11.

SEDGEWICK, R.; FLAJOLET, P. *An Introduction to the Analysis of Algorithms*. Pearson Education, 2013. ISBN 9780133373486. Disponível em: <<https://books.google.com.br/books?id=P3tCB8Q7mA8C>>. Citado na página 12.

STOCKMAN, A. K.; BEAMER, D. A.; BOND, J. E. An evaluation of a garp model as an approach to predicting the spatial distribution of non-vagile invertebrate species. *Diversity and Distributions*, v. 12, n. 1, p. 81–89, 2006. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1366-9516.2006.00225.x>>. Citado na página 12.

WISZ, M. S. et al. Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, v. 14, n. 5, p. 763–773, 2008. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1472-4642.2008.00482.x>>. Citado na página 12.

WOOD, S. *Package 'mgcv'*. 1.9-3. ed. [S.l.], 2025. <https://CRAN.R-project.org/package=mgcv>. Disponível em: <<https://cran.r-project.org/web/packages/mgcv/mgcv.pdf>>. Citado na página 19.

## Apêndices

## APÊNDICE A – Exemplo de seção de anexo

EXEMPLO DE CODIGO A SER ADICIONADO