

# Data Science em soluções de Engenharia e Infraestrutura

Como a inteligência de dados contribui para soluções de engenharia



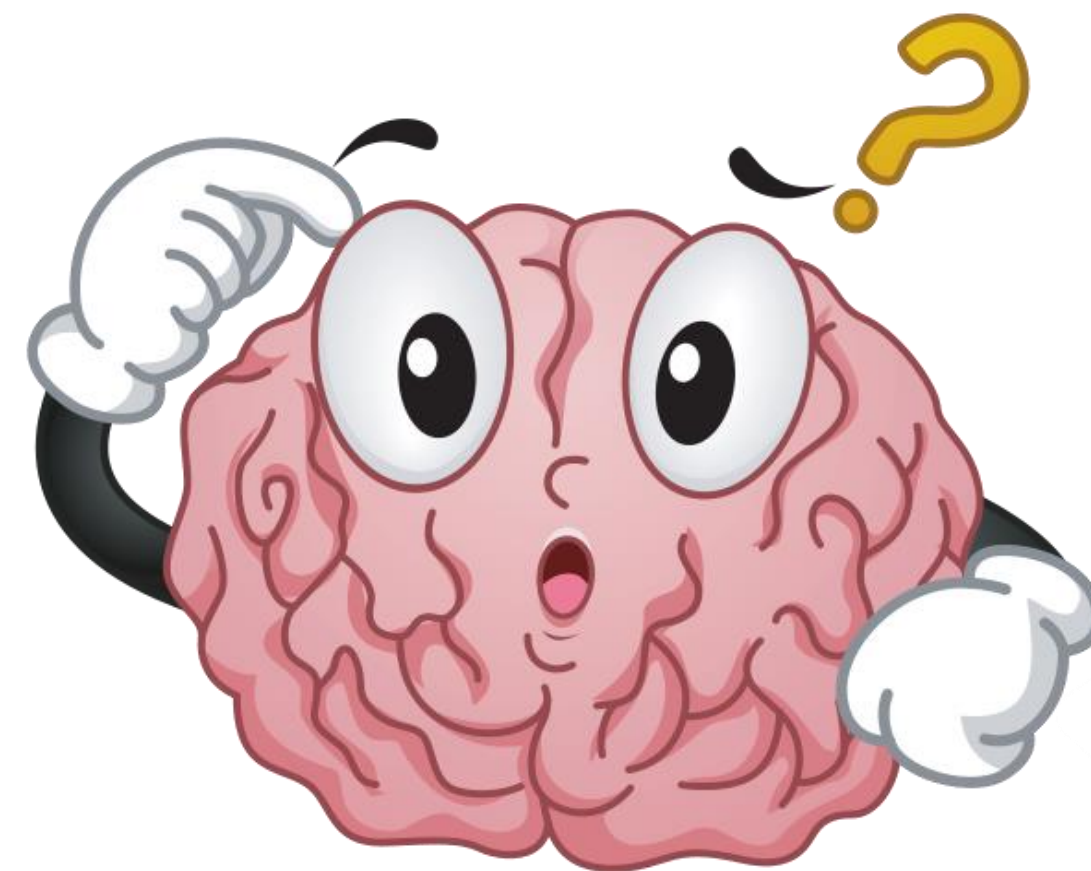
Tiago Marum  
Diretor Técnico  
THM Estatística

Data: 17 de abril de 2024  
Hora: 19h

[www.thmestatistica.com.br](http://www.thmestatistica.com.br)

# Lembretes

A palestra é pra você que está aqui!  
Então, seguem algumas sugestões da THM para uma melhor experiência:



- 1 Não deixe a apresentação avançar com **dúvidas!** A ideia da conversa é **discutir, aprender e esclarecer**
- 2 Sinta-se à vontade para digitar suas perguntas no chat ou abrir o microfone e perguntar
- 3 Algumas questões podem ser respondidas fora do horário da palestra (dependendo do escopo) - deixo meu contato no final



# Agenda

## Sobre o que falaremos esta noite

- O que é Data Science/Machine Learning/Estatística Aplicada?
- Quais técnicas e pra que servem
- Aplicações em Cases de Engenharia
- Por onde começar uma boa análise de dados?
- Que ferramentas são necessárias para criar soluções em dados?





Sobre os palestrantes

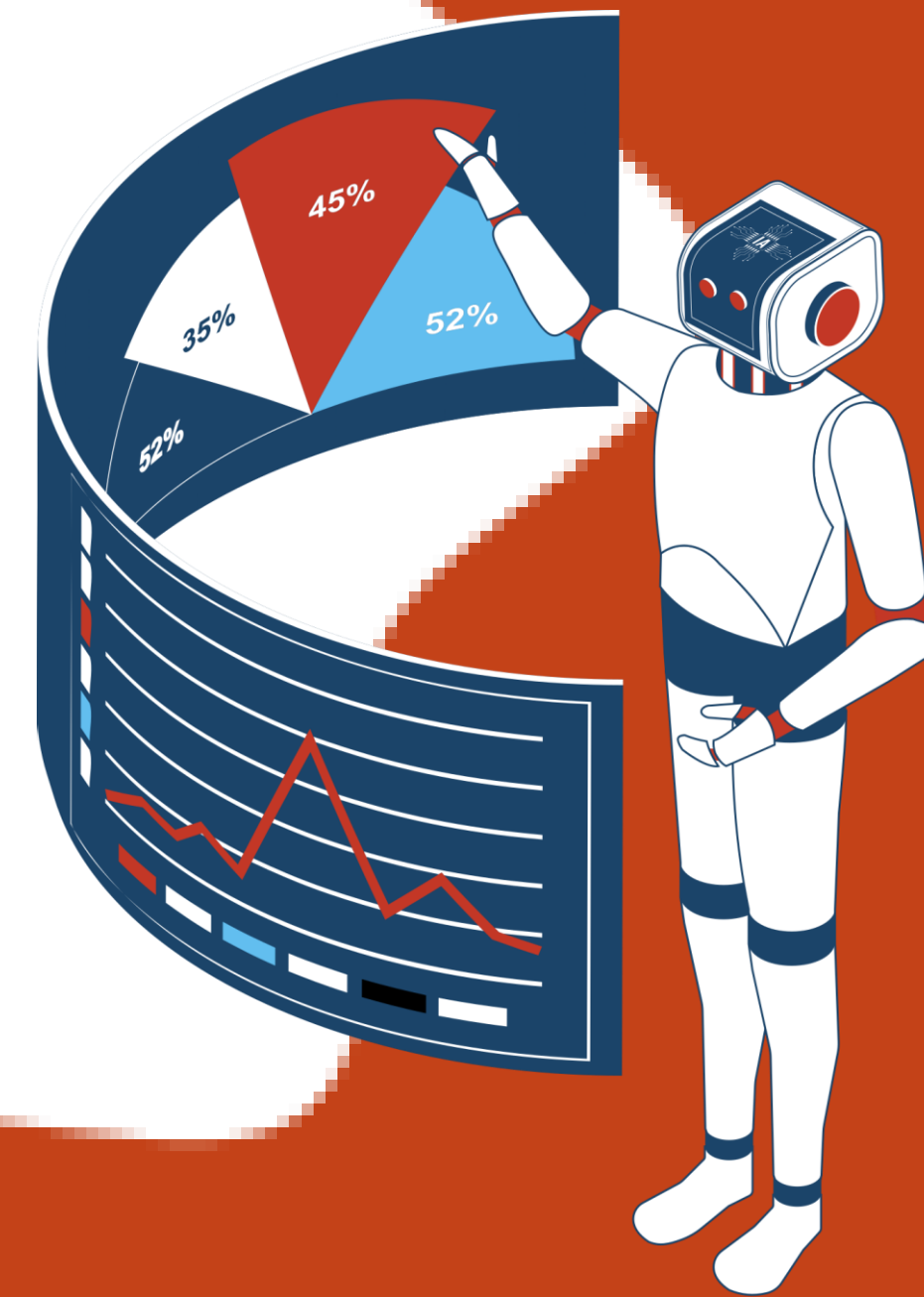
## Somos Tiago Marum e Ronney Agra, da THM Consultoria Estatística

**A THM é uma empresa de consultoria e educação que oferece soluções em dados para o mundo dos negócios e para o meio acadêmico-científico.**

Mestres e Engenheiros Civis com formação pela Universidade de São Paulo (USP) e Universidade Federal de Alagoas (UFAL) com expertise em estatística, análises geoespaciais e desenvolvimento de sistemas. Além disso, somos professores e autores de conteúdo acadêmico e material didático em engenharia, probabilidade e estatística.

# O que é Data Science?

É uma análise? É um programa?  
É um termo da moda?

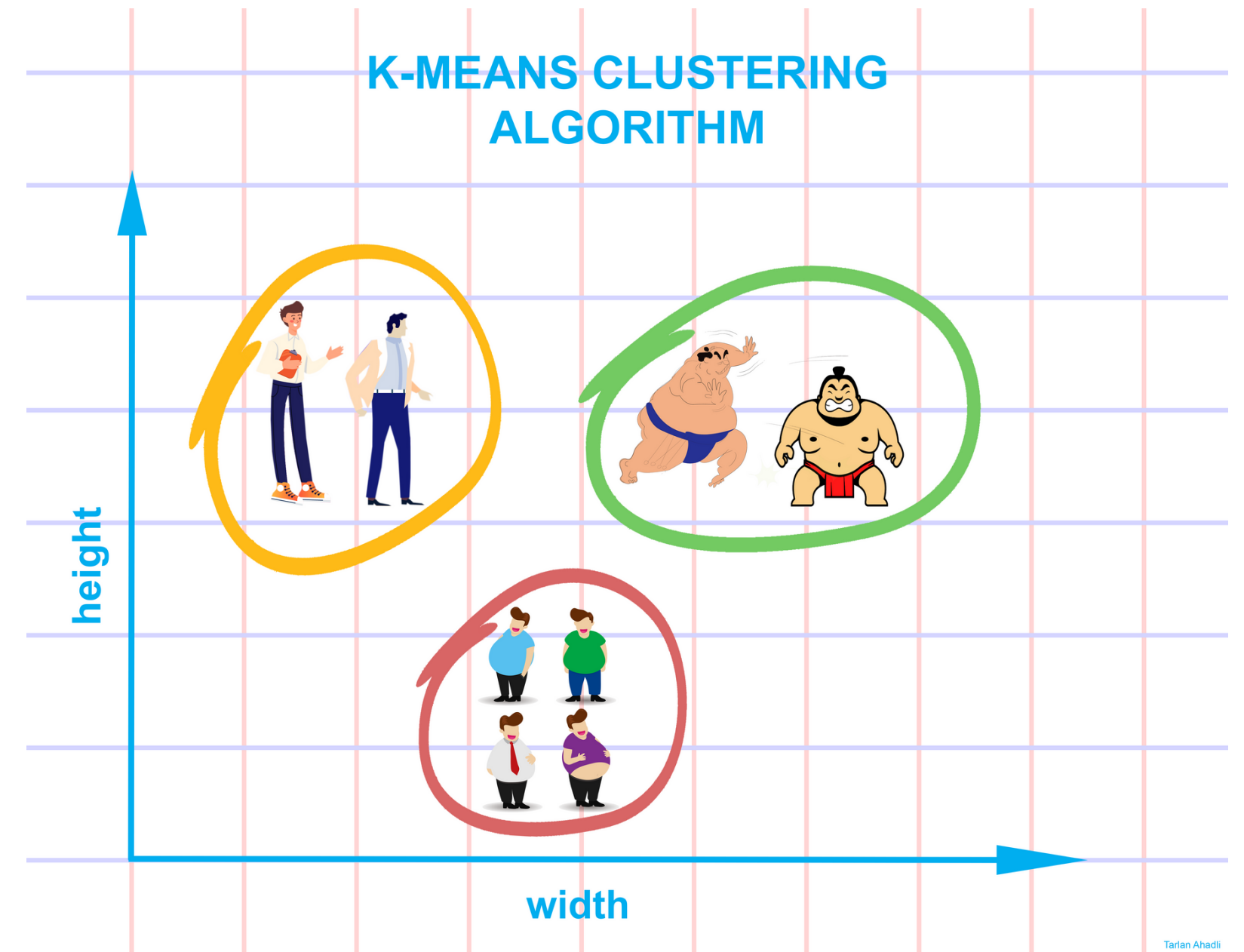


## Definindo Data Science

É um campo interdisciplinar que envolve aplicação de métodos estatísticos, algoritmos de Machine Learning e técnicas de visualização.

Aplicando esses “ingredientes” é possível extrair insights de conjuntos de dados complexos. A ciência de dados envolve todas as etapas do processo de análise de dados, desde a coleta e limpeza de dados, análise exploratória até a construção de modelos preditivos

Exemplo de técnica para reconhecimento de padrões (clusters) nos dados

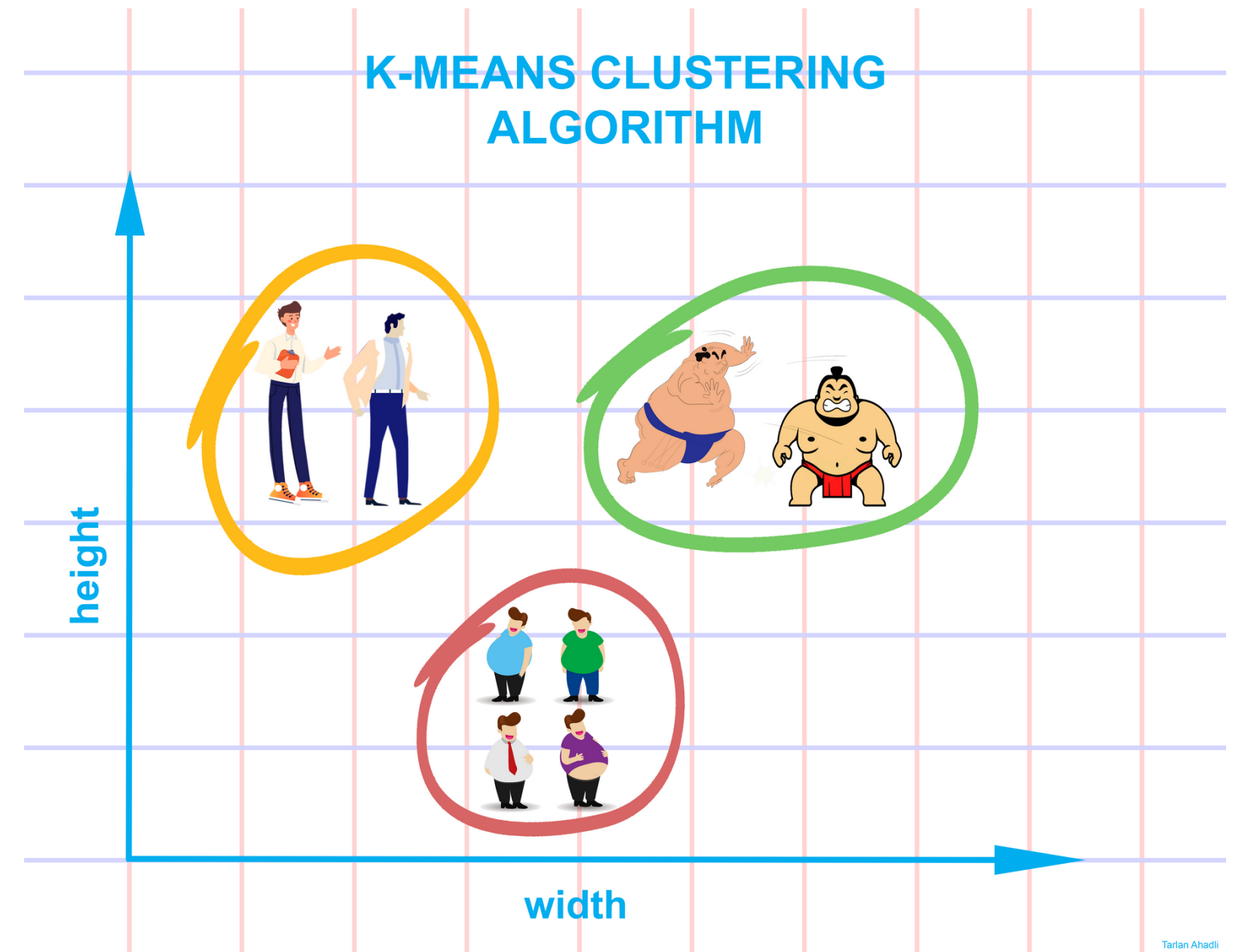


## Definindo Data Science

É  
q  
n  
a  
J  
ts  
s

Tá, mas o que são  
então esses  
“métodos  
estatísticos e  
algoritmos de  
Machine Learning?”

Exemplo de técnica para reconhecimento  
de padrões (clusters) nos dados



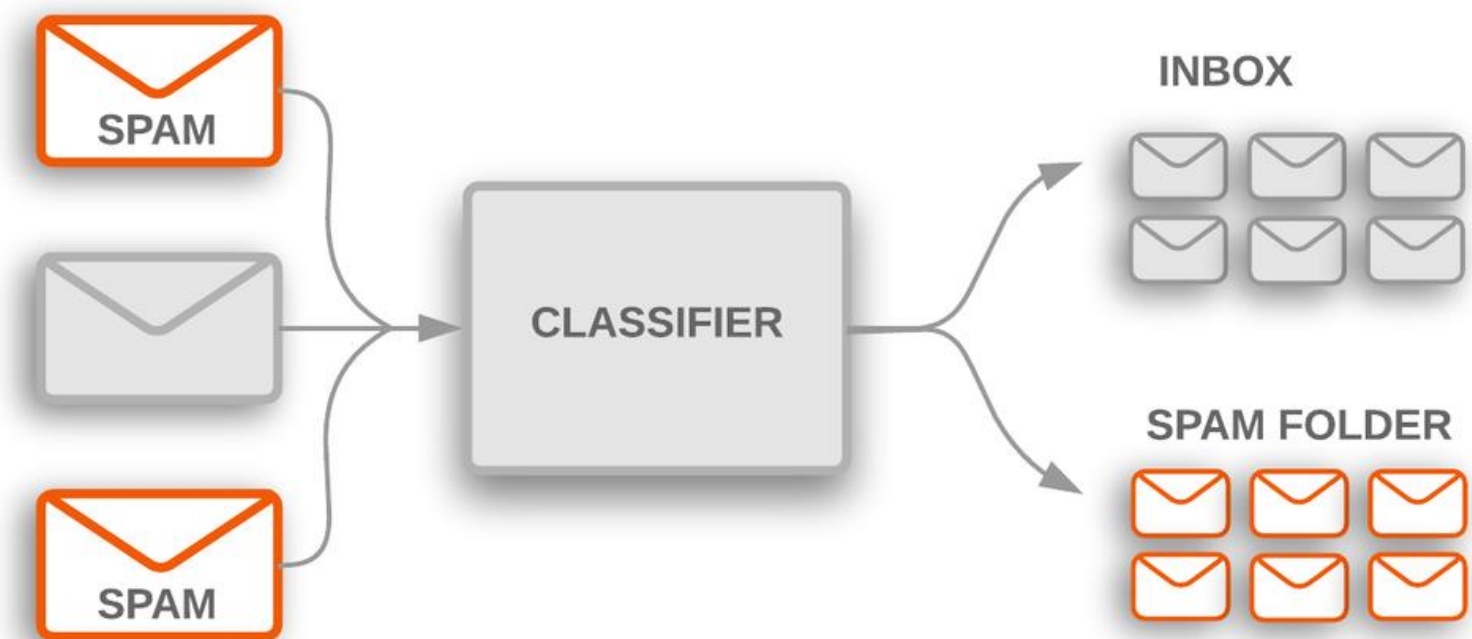
## Definindo Machine Learning

É um subcampo da ciência de dados focado no desenvolvimento de algoritmos e modelos que podem aprender padrões e fazer previsões a partir de dados

Se baseia em treinar modelos em conjuntos de dados históricos e usar esses modelos para fazer previsões ou tomar decisões em novos dados. Exemplo: reconhecimento de padrões, processamento de linguagem natural, visão computacional, otimização de processos...

Exemplo de conceitos do algoritmo Naive-Bayes, muito utilizado para classificação de texto ou documentos envolvendo conjuntos com alta dimensão

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B)}$$



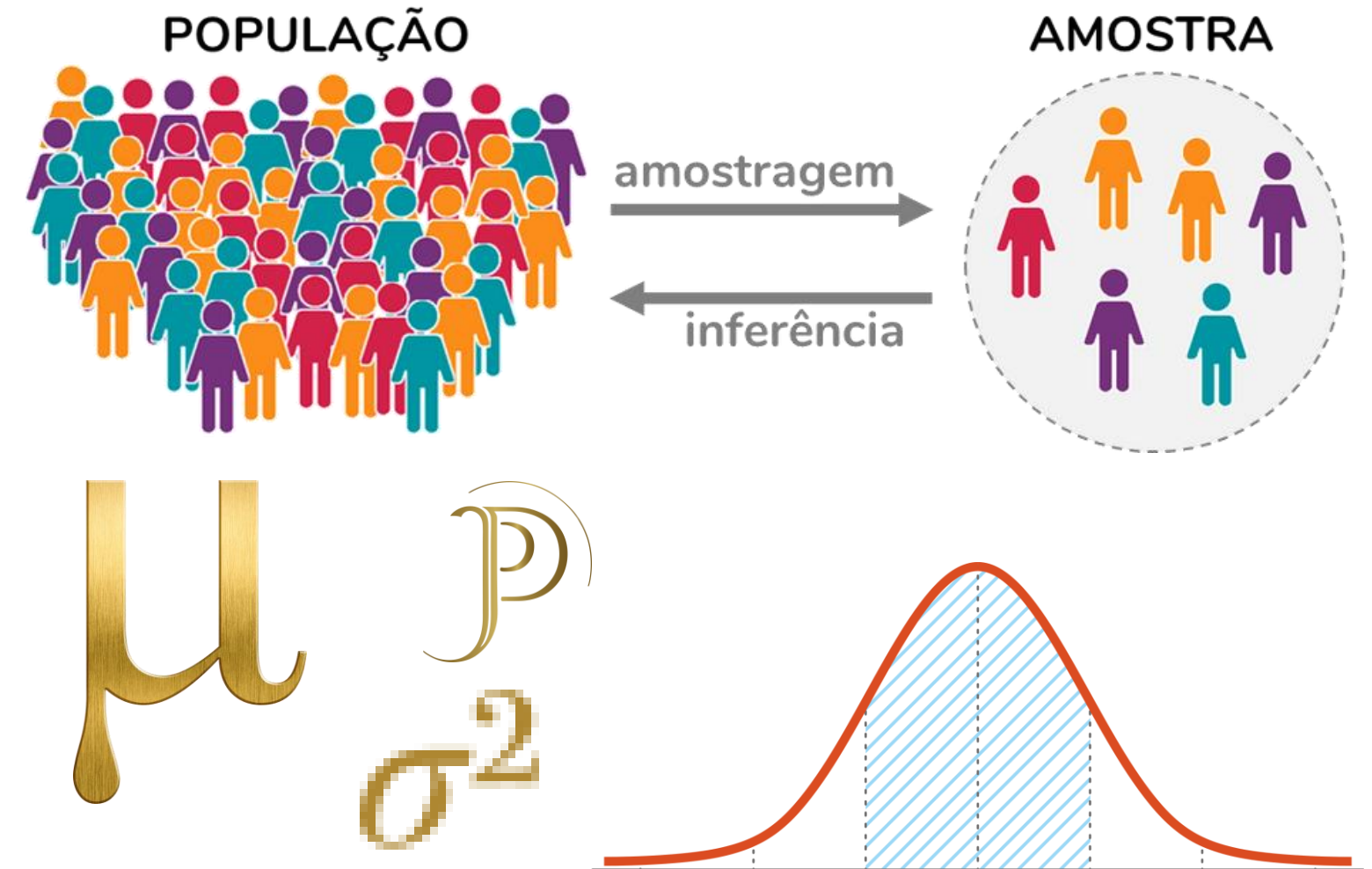


## Definindo Estatística

É o estudo de como descrever e analisar dados para fazer inferências sobre populações maiores. Fundamental para análise quantitativa e tomada de decisões baseadas em evidências

Métodos estatísticos são frequentemente utilizados para calcular probabilidades, estimar parâmetros populacionais (como média, desvio padrão, coeficiente de variação, proporções, máximos, mínimos), testar hipóteses e modelar relações entre variáveis

Conceitos de inferência: população x amostra, parâmetros estatísticos e distribuições probabilísticas

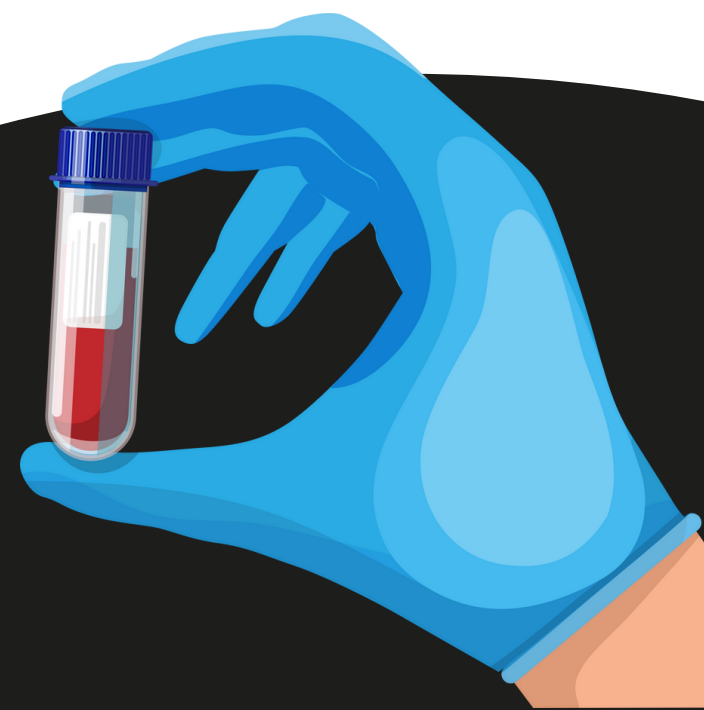


# Tudo começa com Inferência Estatística!

Para aproveitar o melhor da Ciência de Dados é necessária uma forte base em Inferência, que depende principalmente de três conceitos:



**Teoria de Probabilidades**

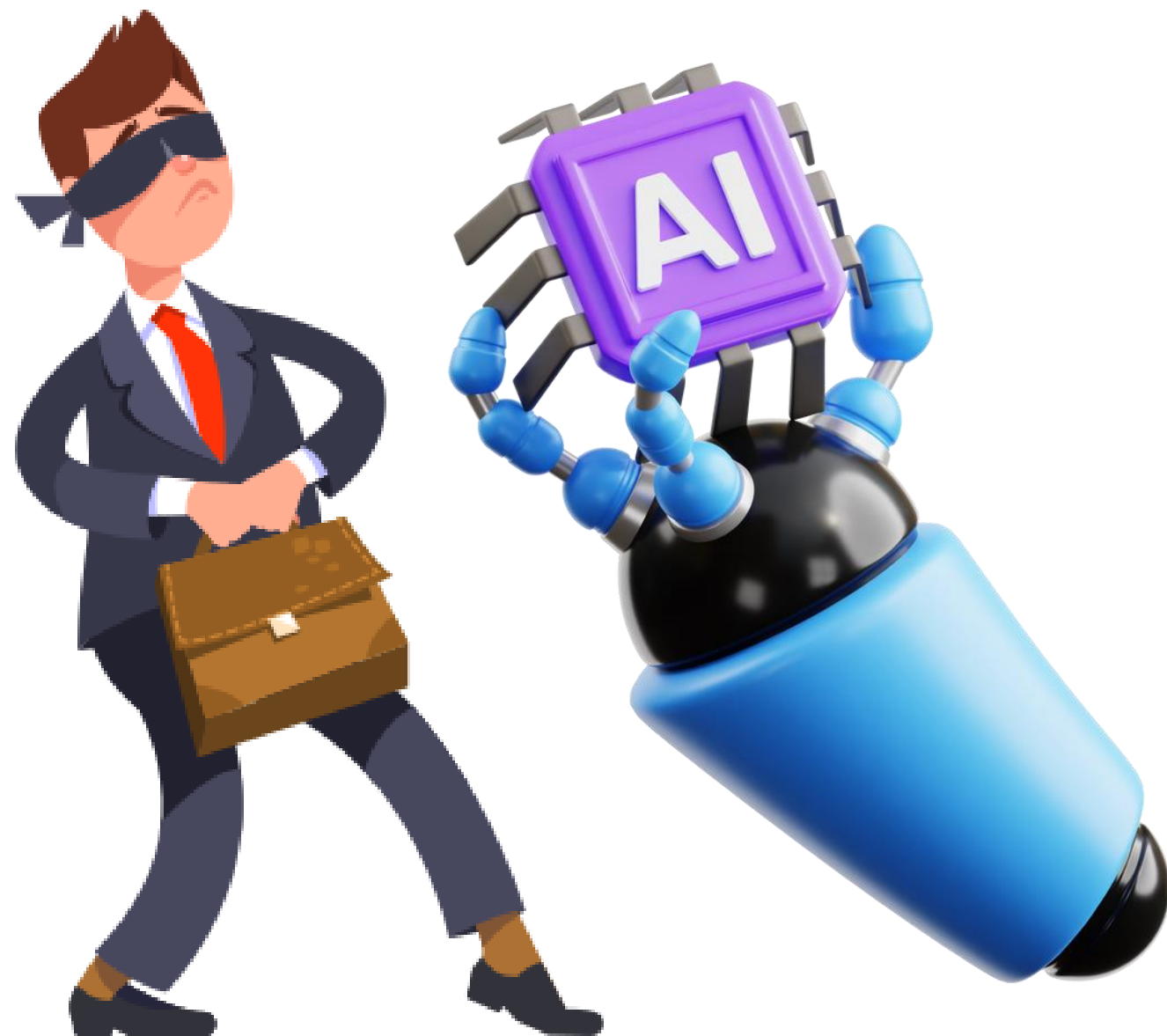


**Conceitos de Amostragem**



**Análise Descritiva de Dados**

Mais comum do  
que você pensa



# 5 em 5 engenheiros

tiveram algum contato com  
probabilidade e inferência estatística  
durante a formação

Infelizmente, a falta de “colocar a teoria em prática” ainda deixa a desejar, fazendo com que o futuro engenheiro não consiga enxergar a verdadeira capacidade das ferramentas que são ensinadas. No mercado, um engenheiro com habilidade de tomar decisões baseadas em dados se diferencia pela assertividade e repertório! Grandes gestores e líderes consideram abordagens Data-Driven.



Educacional THM

Ordenar

-  **Introdução | Curso de Probabilidade 01**  
Educacional THM • 8,5 mil visualizações • há 2 anos  
13:24
-  **Probabilidade, Espaço Amostral e Diagrama de Venn | C**  
Educacional THM • 3 mil visualizações • há 2 anos  
21:33
-  **Probabilidade Condicional | Curso de Probabilidade 03**  
Educacional THM • 2,4 mil visualizações • há 2 anos  
22:20
-  **Teorema de Bayes | Curso de Probabilidade 04**  
Educacional THM • 2 mil visualizações • há 2 anos  
21:50

# Por onde começar?

Playlists completas de Probabilidade e Estatística no YouTube no Canal Educacional THM.

O canal é Top3 no Youtube em recomendações orgânicas nas pesquisas por “Curso de Probabilidade” e “Curso de Estatística”.

# Quais são e onde aplicar todas essas ferramentas?

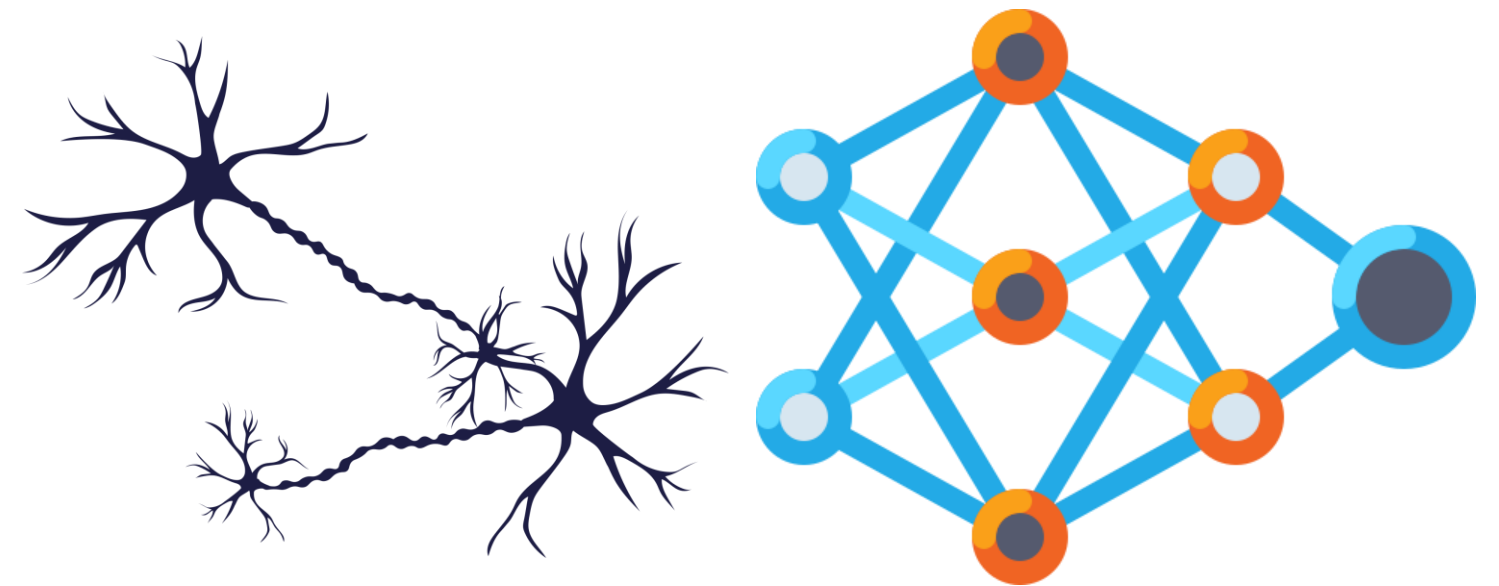
Exemplos e discussões



## Redes Neurais Artificiais

Modelos computacionais inspirados no funcionamento do cérebro humano. Consistem em camadas de neurônios interconectados e são capazes de aprender e generalizar padrões complexos a partir de grandes conjuntos de dados.

Exemplo esquemático de funcionamento de redes neurais



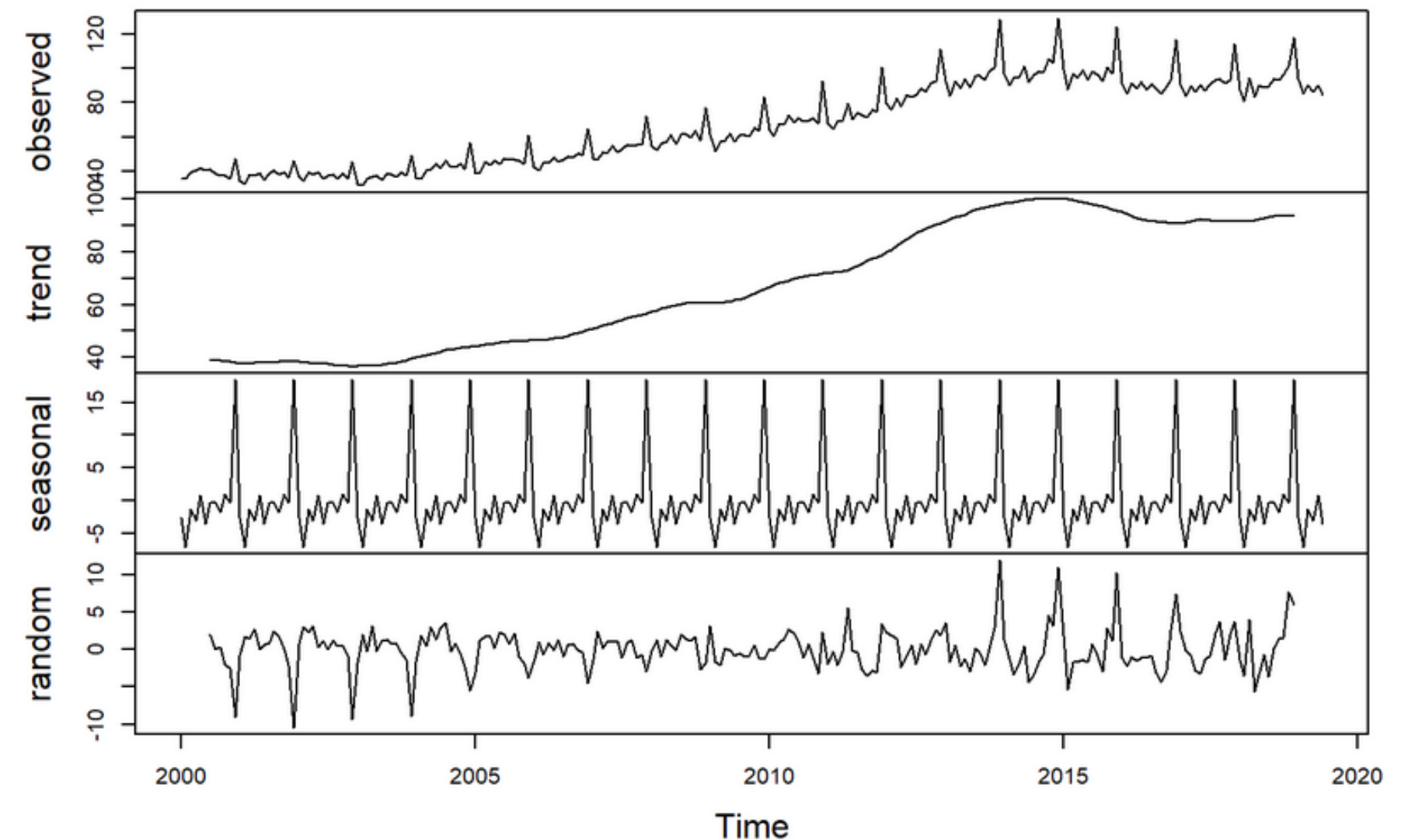
- Processamento de Sinais e Imagens
- Controle de Processos
- Sistemas Embarcados e IoT
- Otimização e Projeto de Estruturas
- Processamento de Linguagem Natural (PLN)

## Modelagem de Séries Temporais

As técnicas de séries temporais visam analisar e prever padrões em dados que variam ao longo do tempo. São amplamente utilizadas em previsão financeira, previsão de demanda, análise climática e muitas outras aplicações.

Decomposição de série temporal: Volume do Varejo Total em Mato-Grosso do Sul (2010-2019)

Decomposition of additive time series

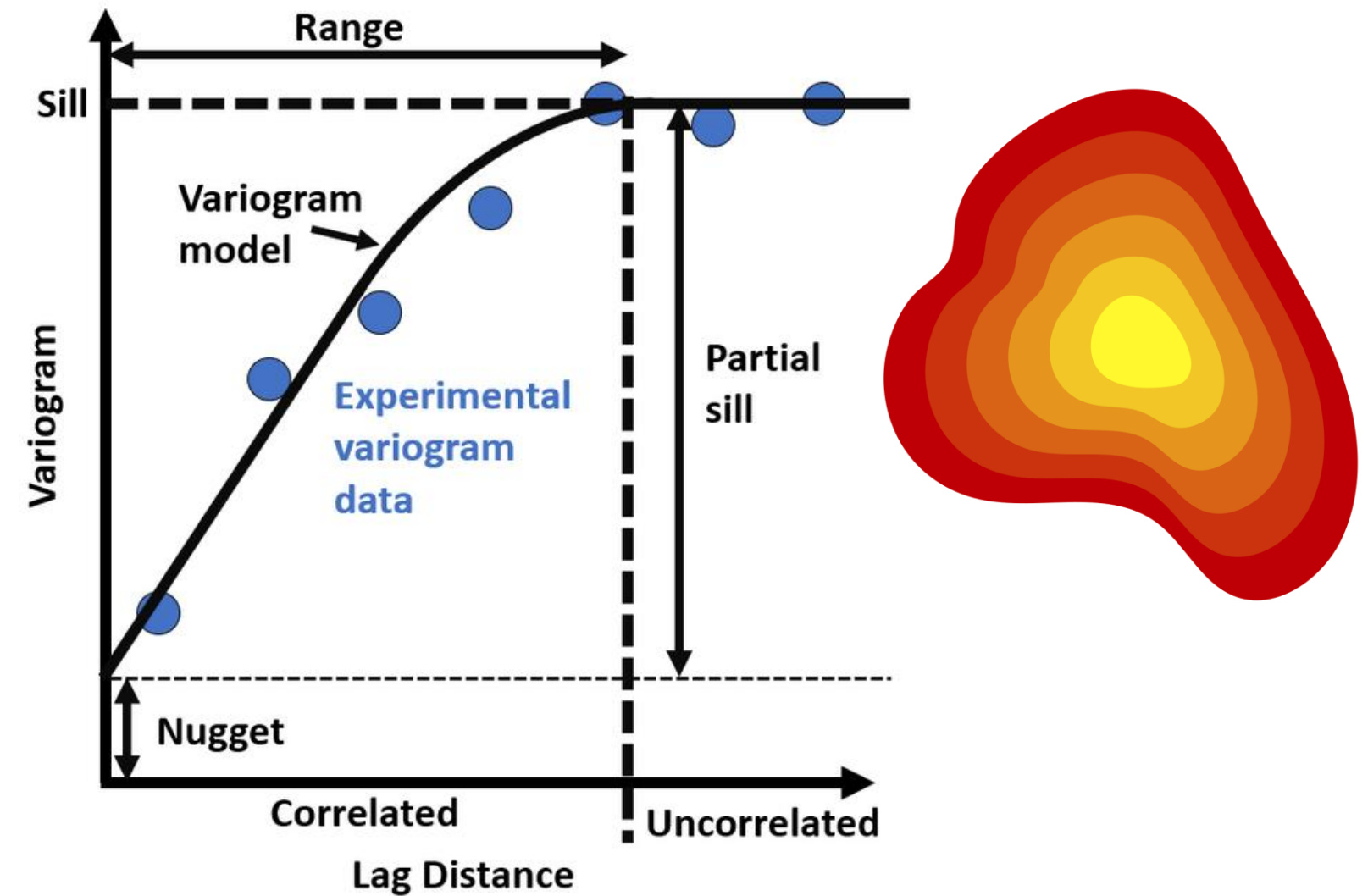


A escolha da técnica adequada depende da natureza dos dados, da complexidade dos padrões temporais e dos objetivos da análise.

## Análises Geoespaciais

A geoestatística é uma ferramenta poderosa para analisar e modelar dados geográficos. As técnicas de geoestatística permitem não apenas interpolar dados, mas também quantificar e lidar com incertezas espaciais de forma rigorosa.

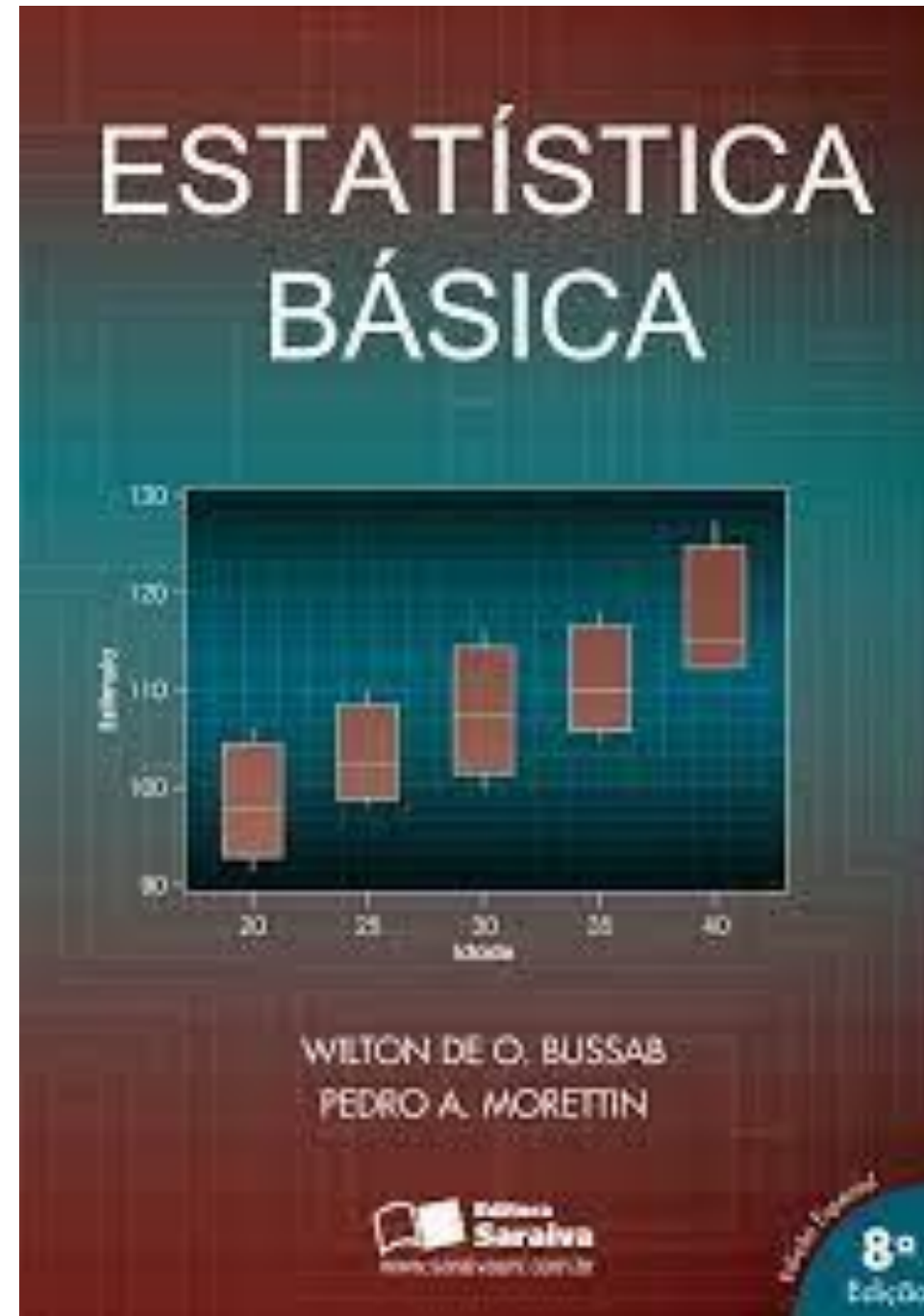
Exemplos de variograma e superfície interpolada por métodos geoestatísticos



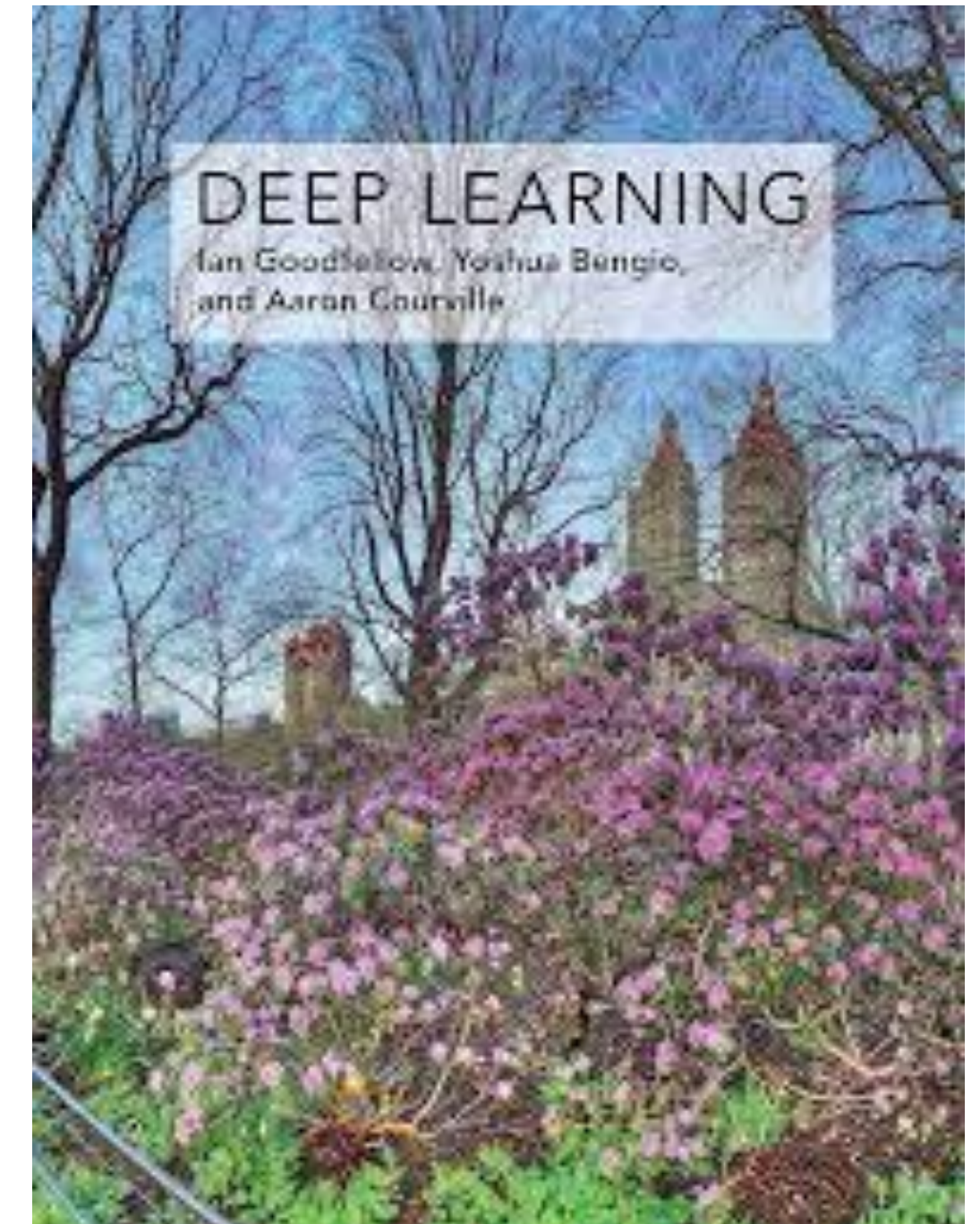
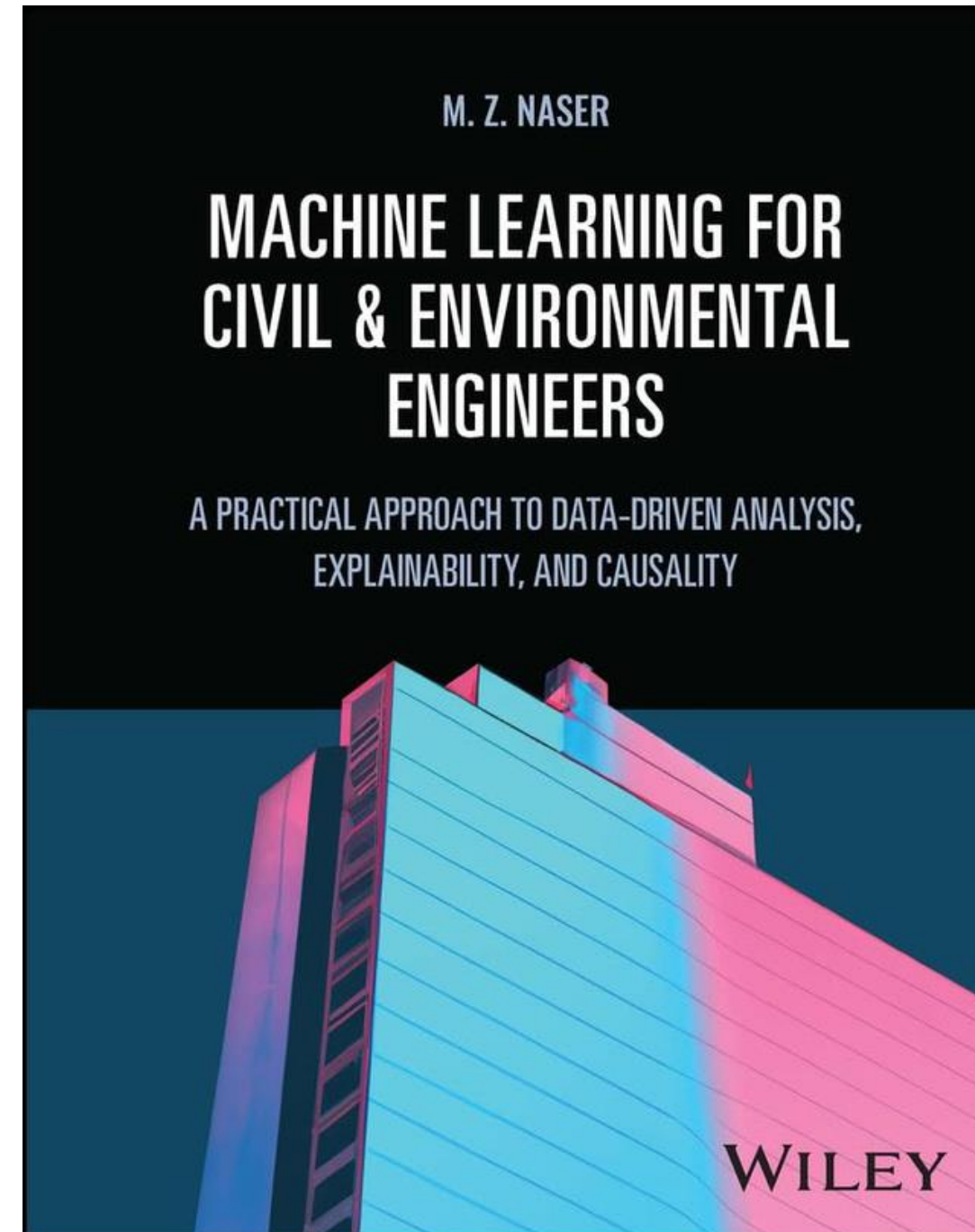
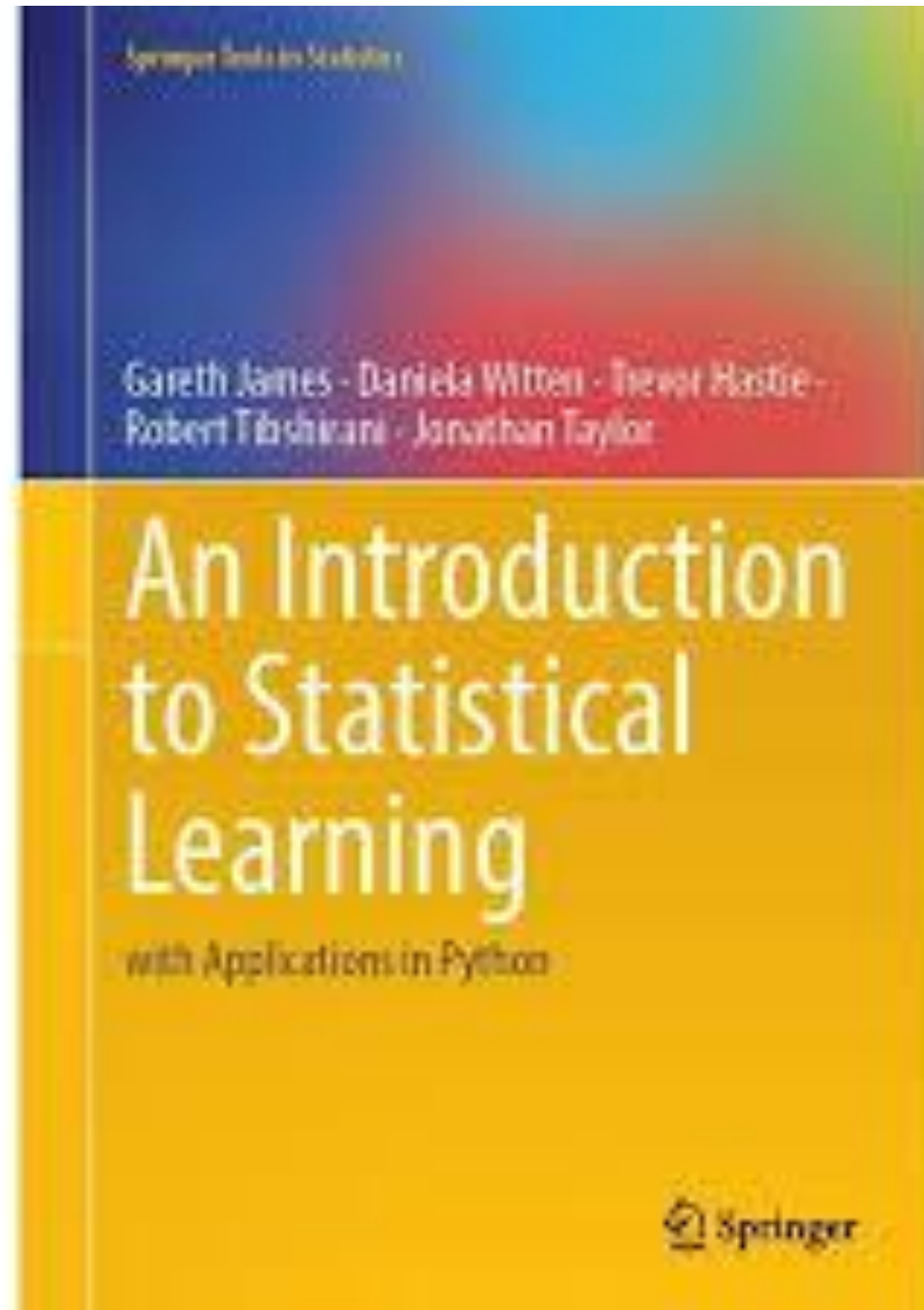
Usada especialmente quando há uma estrutura espacial nos dados. para entender a distribuição e a variabilidade espacial dos fenômenos estudados, fazer previsões e tomar decisões informadas.



## Referências



## Referências



# Como sei qual técnica devo usar?



Domínio de Fundamentos, repertório, experiência.

Em inteligência de dados você não tem um só caminho para solucionar um problema (engenheiros capacitados vão saber tomar decisões frente a isso).



# A conversa está longe de acabar por aqui

O mundo de dados é vasto e não conseguirá ser resumido a uma conversa de 2h.

Alguns exemplos:

- Engenharia Civil e Produção: Análises TBM, Construção Civil (acompanhamento, monitoramento de processos), termografia embarcada em drone
- Engenharia de Minas, Ambiental: Gerenciamento de Áreas Contaminadas (análises multivariadas)
- Engenharia Mecânica e Civil: Sensor de Irregularidade do Pavimento
- Engenharia Elétrica e Smart Cities: Sensores de Iluminação Pública (medição de lux)
- Multidisciplinar: A THM em Mariana e Brumadinho
- Engenharia Ambiental, Química: Análises de concentrações químicas em Água Superficial (linguagem R, faxina de dados e discussão prática)



# Obrigado!

Envie uma mensagem  
no email [marum@thmestatistica.com](mailto:marum@thmestatistica.com)  
se tiver alguma pergunta.

