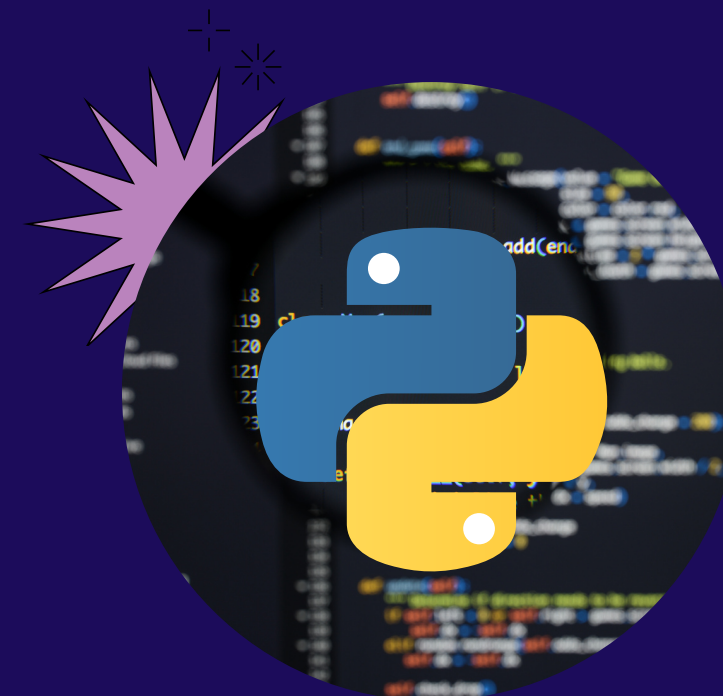


# PYTHON – NÍVEL INICIANTE

*POWERED BY MULHERES EM DADOS*

## PROJETO SATISFAÇÃO DO CONSUMIDOR



AULA 03 – FEATURE ENGINEERING E CRIAÇÃO DE MODELO DE MACHINE  
LEARNING

29 DE JUNHO DE 2022

# CRONOGRAMA

JUNHO-JULHO 2022

Acesse [aqui](#)  
o notebook  
da aula 03

DOM	SEG	TER	QUA	QUI	SEX	SAB
12	13	14	15 Entendimento	16	17	18
19	20	21	22 Análise Exploratoria	23	24	25
26	27	28	29 Criação de modelo	30	1	2
3	4	5	6 Avaliação	7	8	9

## Entendimento, carregamento e pré-processamento dos dados

Entendimento do negócio e configurar o colab para pre-processar os dados

## Análise exploratória de dados (EDA)

Analisaremos os dados resultando nos principais insights sobre o negócio

## Feature Engineering e criação de modelo de ML

Criaremos o nosso modelo de machine learning a partir das variáveis que mais fazem sentido para o negócio

## Avaliação e melhorias no modelo de ML

Entenderemos melhor sobre como avaliar e melhorar a performance do nosso modelo de machine learning

# AULA 03 – FEATURE ENGINEERING E CRIAÇÃO DE MODELO DE MACHINE LEARNING

## SUMÁRIO

1. Engenharia de Características (Feature Engineering)
  - 1.1. Modelo
2. O que é Aprendizado de Máquina?
  - 2.1. Dados de treino e teste
  - 2.2. Árvore de Decisão (Decision Tree)
  - 2.3. Florestas Aleatórias (Random Forests)
3. Referências
4. Para saber mais



# ENGENHARIA DE CARACTERÍSTICAS (FEATURE ENGINEERING)

Tem como objetivo melhorar o desempenho do modelo de machine learning, uma vez que verifica quais dados de entrada são relevantes e compatíveis ao tipo de algoritmo a ser usado.

Para tornar os dados mais precisos, o *feature engineering* é usado para criar novas *features*, como novas colunas e variáveis fictícias, que fazem mais sentido para a criação do nosso modelo de machine learning.

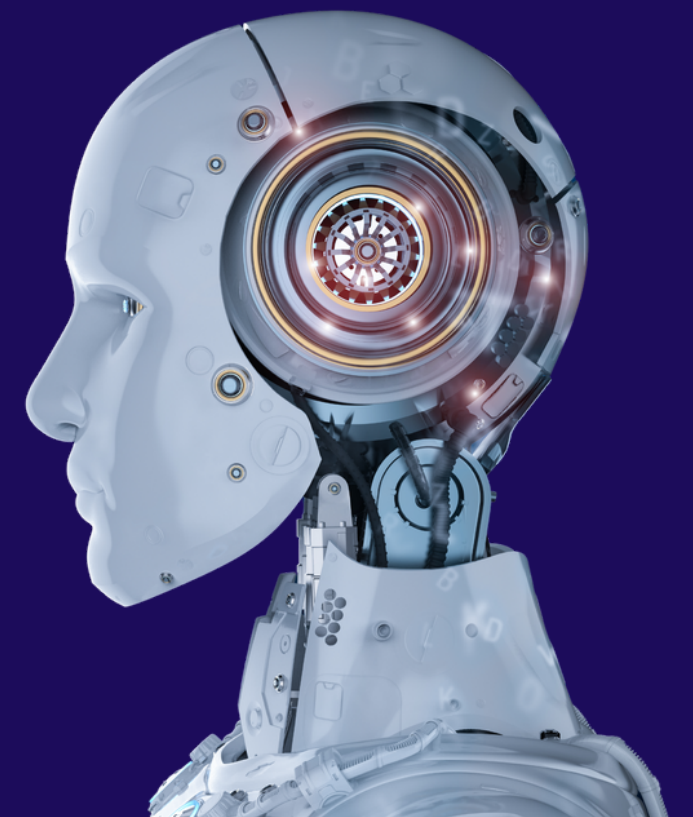
## MODELO

"É uma especificação de uma relação matemática (ou probabilística) entre diferentes variáveis". Temos como exemplo um modelo de negócios, em que "o lucro é a receita menos as despesas, a receita é o número de unidades vendidas multiplicado pelo preço". (GRUS, 2021, p. 157)



# O QUE É APRENDIZADO DE MÁQUINA?

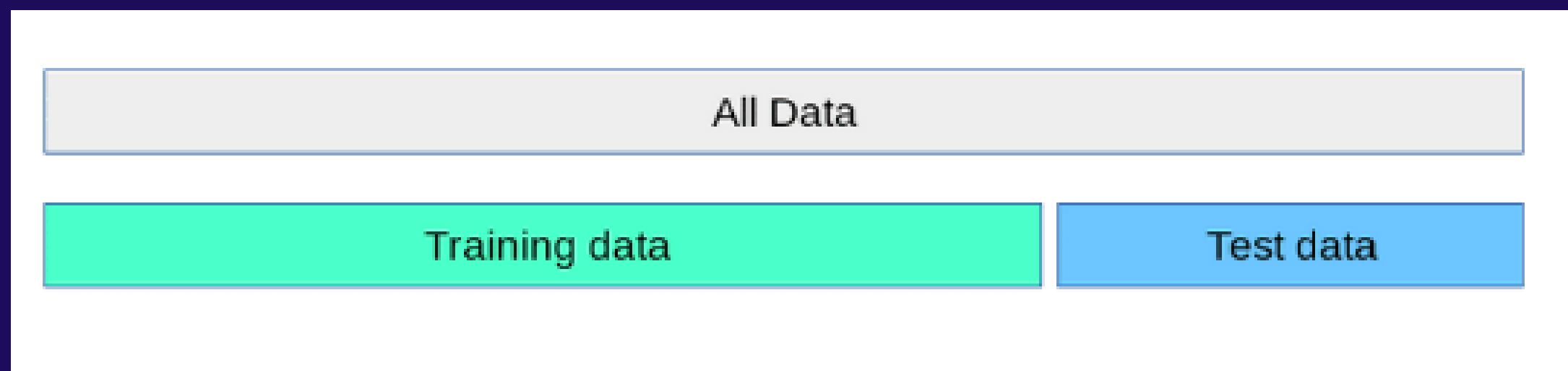
Quando ensinamos máquinas a **aprender padrões observando exemplos de dados (por meio de treinamento e teste)**, de modo que a máquina possa reconhecer esses padrões e aplicar a novos exemplos que ela não tenha visto antes.



# DADOS DE TREINO E TESTE

- Conjunto de **dados de treino**: ajustar o modelo de aprendizado de máquina.
- Conjunto de **dados de teste**: avaliar o modelo de aprendizado de máquina adequado.

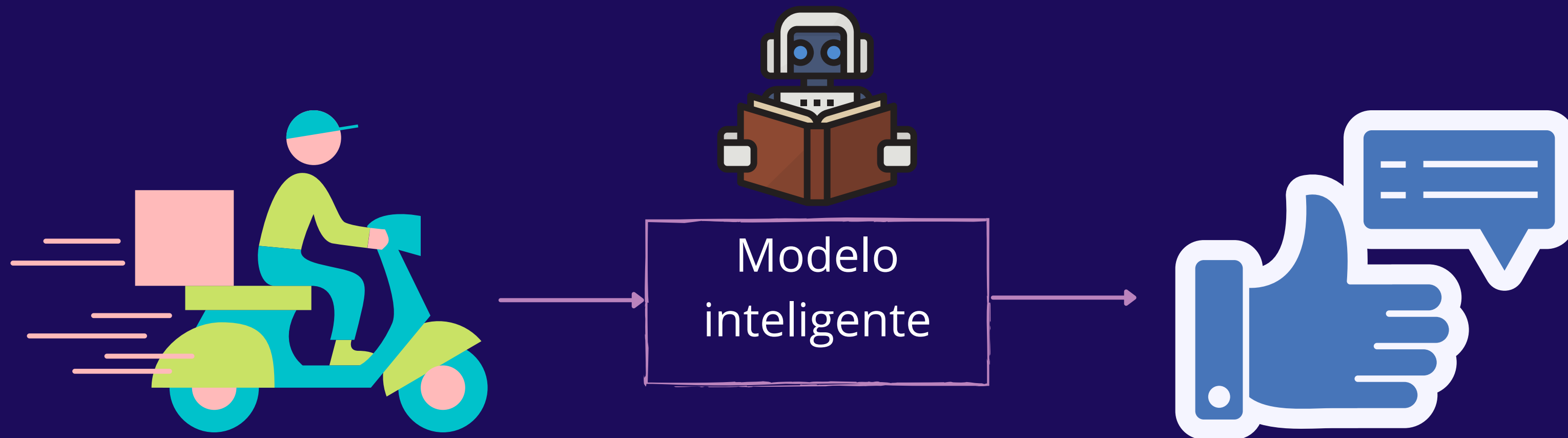
O objetivo é **estimar o desempenho do modelo de aprendizado de máquina** em novos dados: dados não usados para treinar o modelo.





# O QUE É APRENDIZADO DE MÁQUINA?

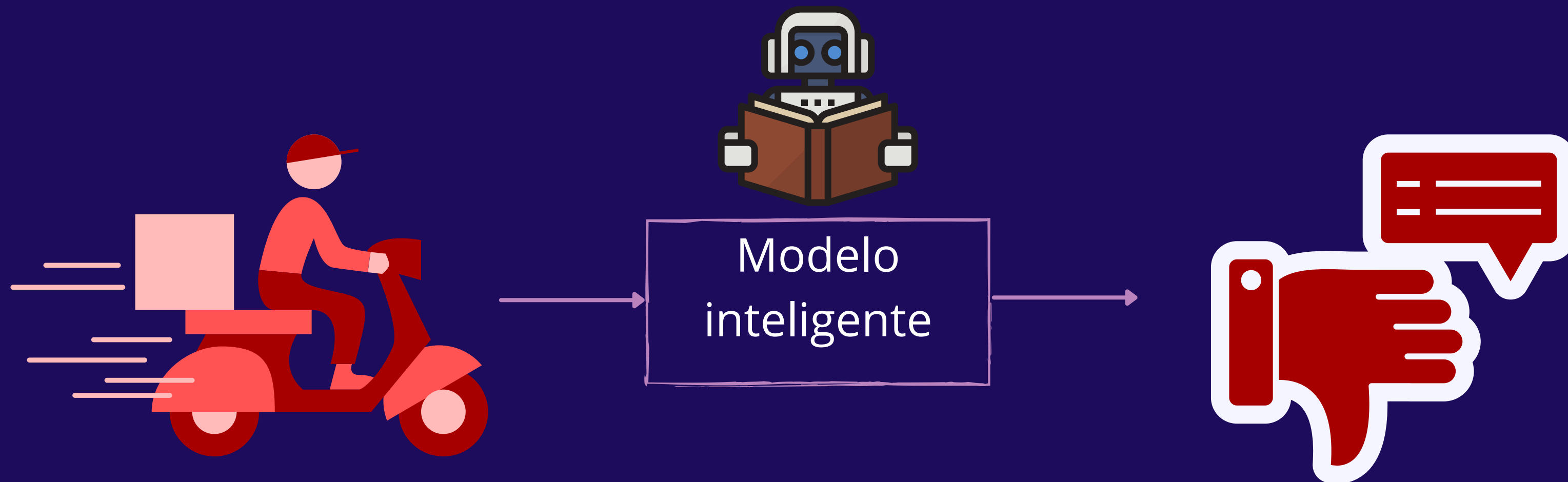
TREINAMENTO:



dias estimados para entrega	dias para a entrega	frete
27	11	16,09

# O QUE É APRENDIZADO DE MÁQUINA?

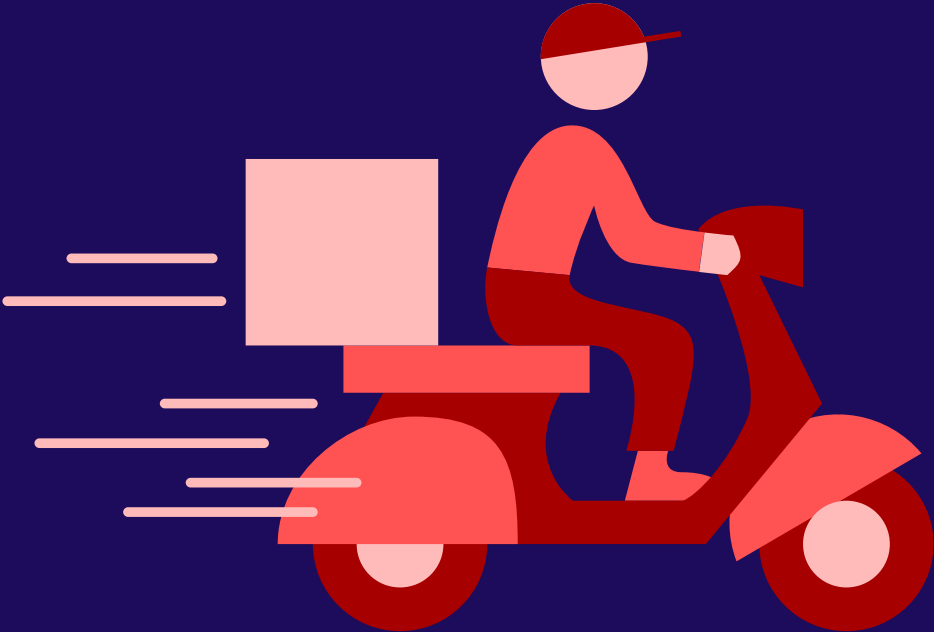
TREINAMENTO:



dias estimados para entrega	dias para a entrega	frete
19	21	36,90



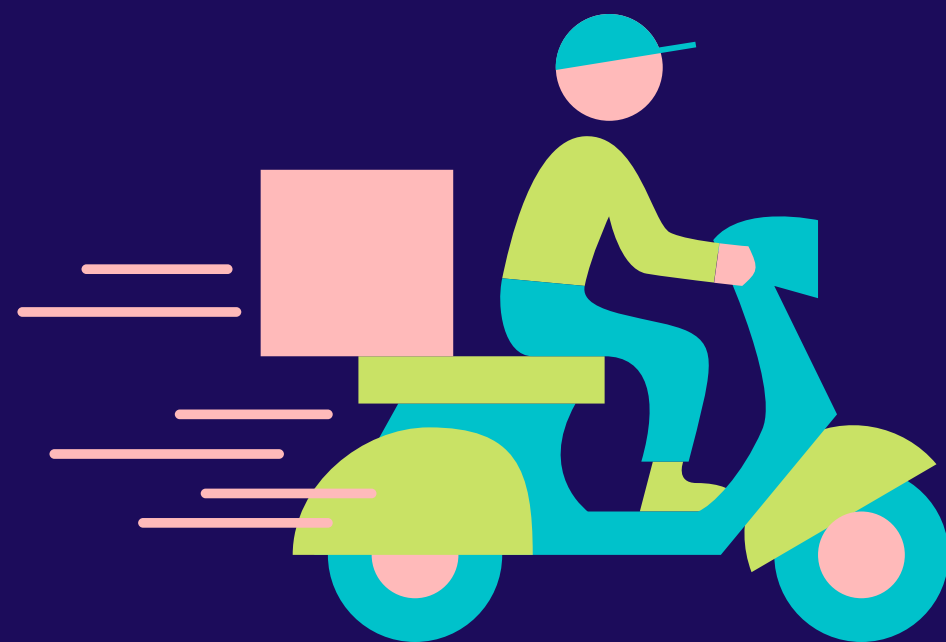
# O QUE É APRENDIZADO DE MÁQUINA?



dias estimados para entrega	dias para a entrega	frete	avaliação
27	11	16,09	positiva
19	21	36,90	negativa

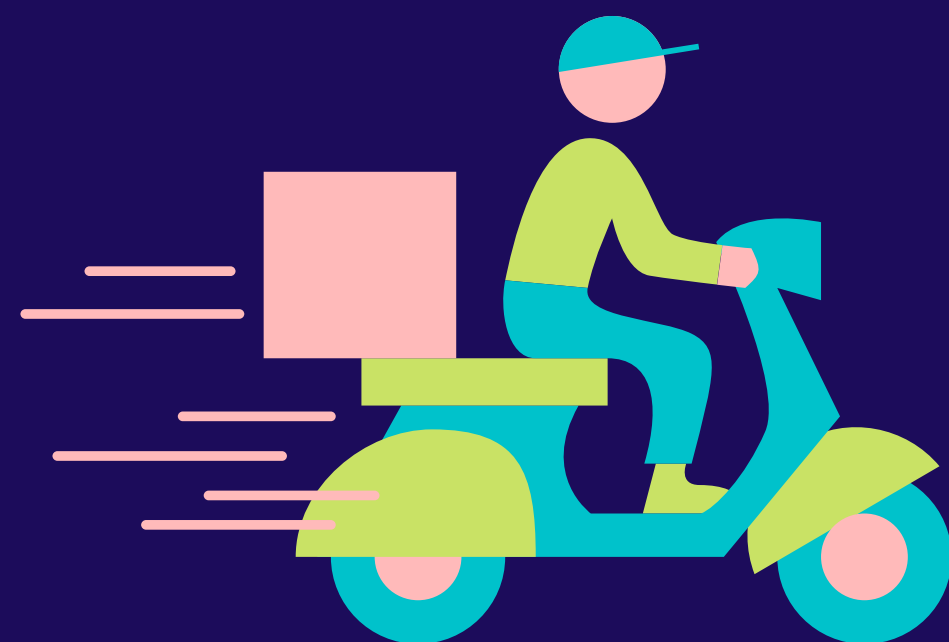


# O QUE É APRENDIZADO DE MÁQUINA?



dias estimados para entrega	dias para entrega a	frete
26	8	18,36

# O QUE É APRENDIZADO DE MÁQUINA?



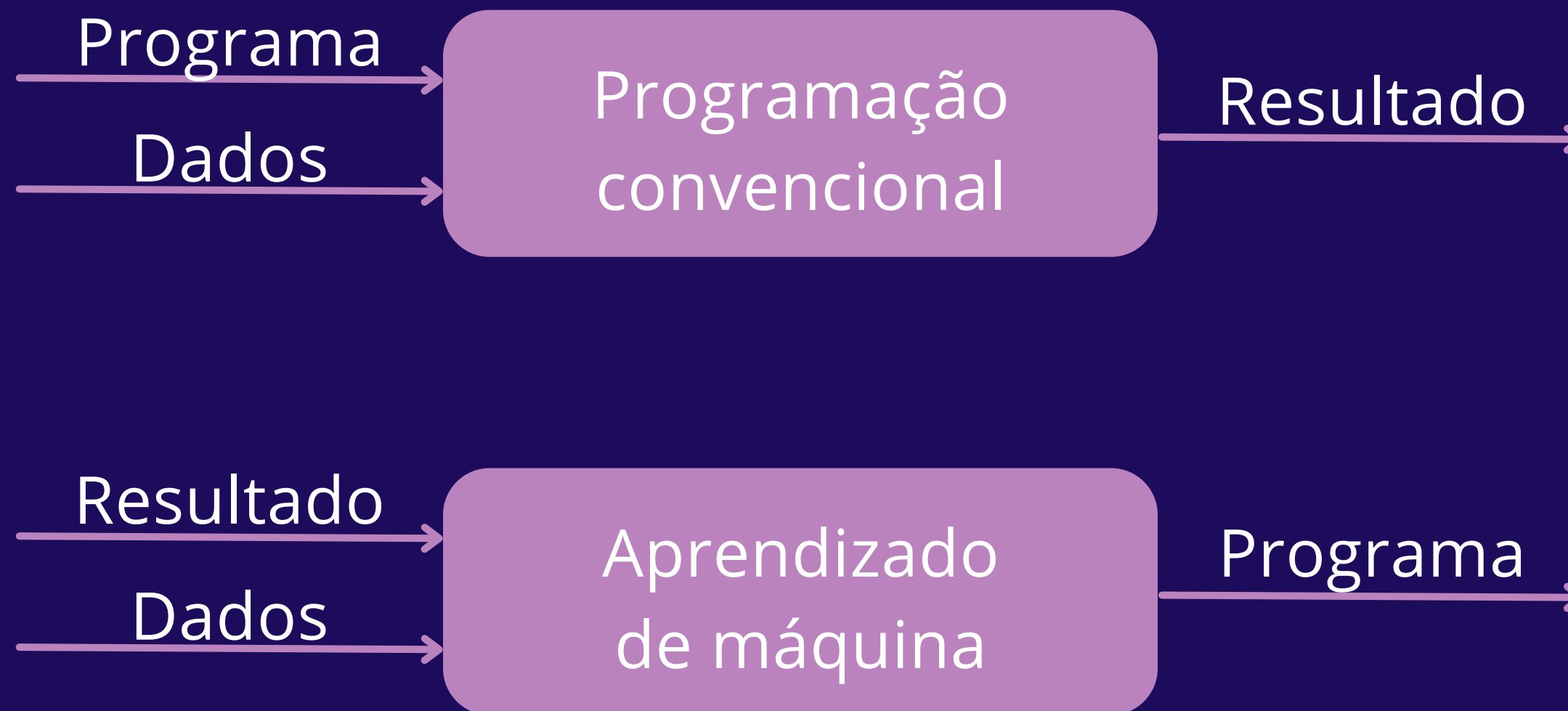
dias estimados para entrega	dias para a entrega	frete
26	8	18,36

# O QUE É APRENDIZADO DE MÁQUINA?



dias estimados para entrega	dias para entrega	frete	avaliação
27	11	16,09	positiva
19	21	36,90	negativa
26	8	18,36	positiva

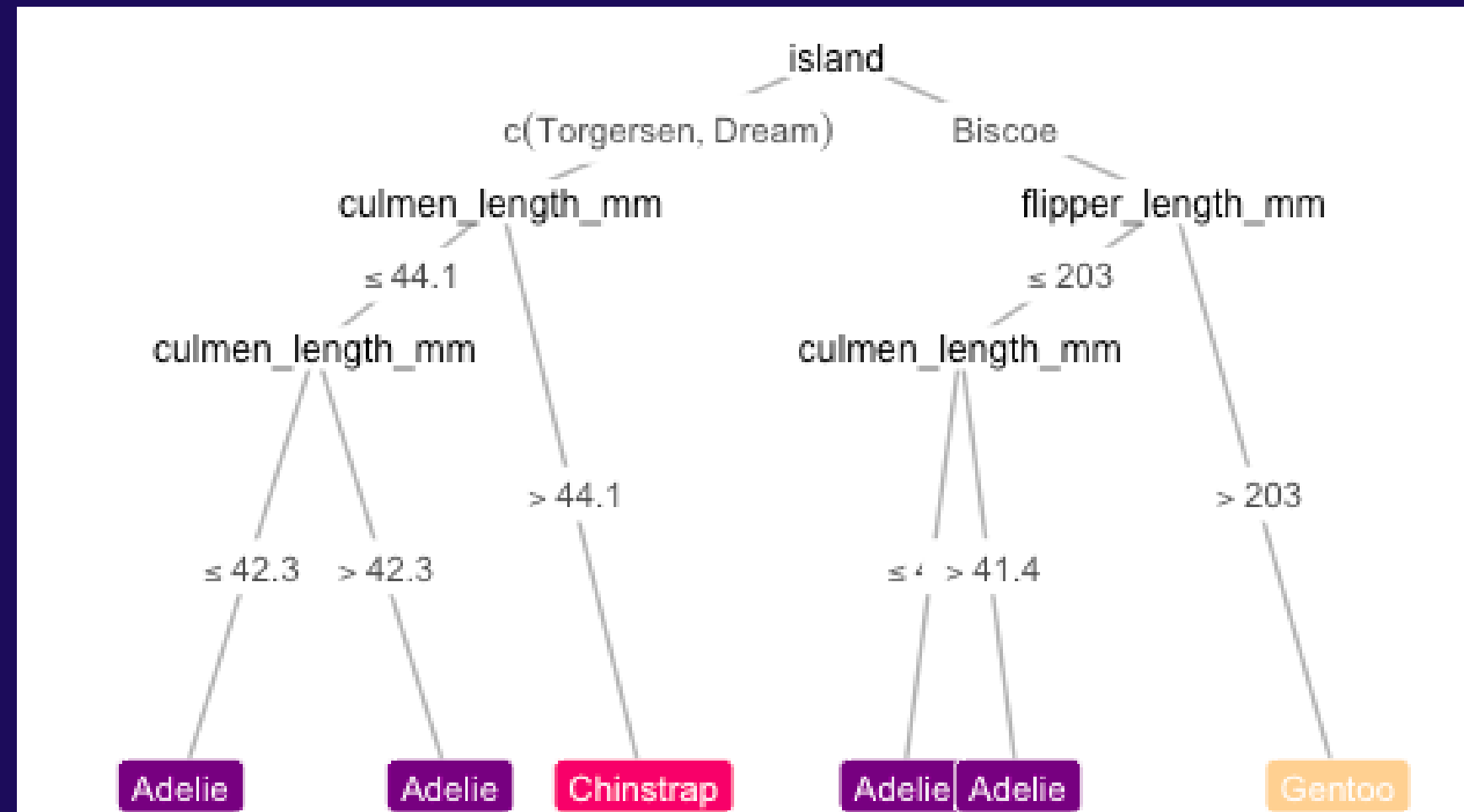
# O QUE É APRENDIZADO DE MÁQUINA?



# ÁRVORE DE DECISÃO (DECISION TREE)

Constroem um **modelo de decisões tomadas com base em valores reais de atributos nos dados**.

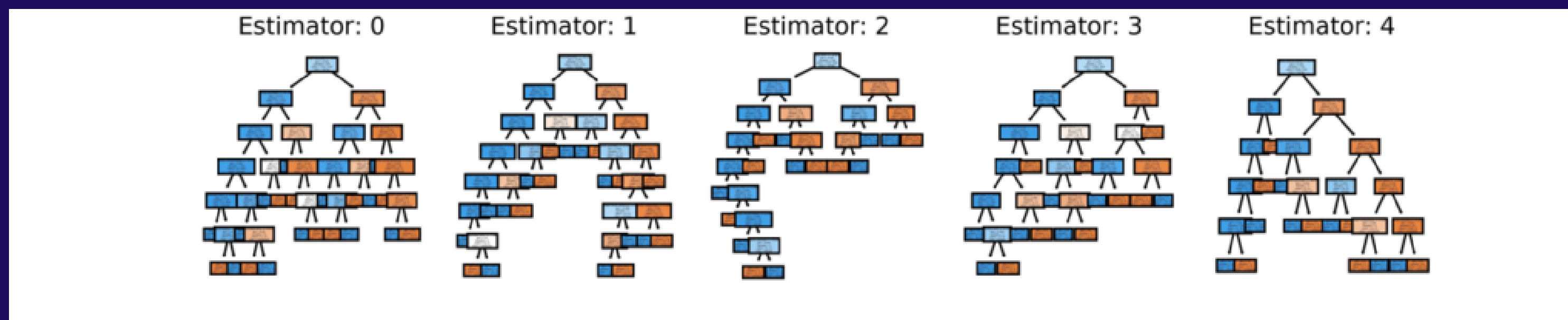
As **decisões se bifurcam** em estruturas de árvore até que uma **decisão de predição** seja tomada.



# FLORESTAS ALEATÓRIAS (RANDOM FORESTS)

Ao mesclar várias árvores de **decisão não correlacionadas**, é possível alcançar uma melhoria significativa da precisão do modelo.

Cada **árvore individual na floresta aleatória retorna uma predição de classe** e a classe com **mais votos** se torna a **predição do modelo**.





# REFERÊNCIAS

(CLIQUE NO LINK)

DATA SCIENCE EU. Engenharia de características.

GRUS, JOEL. Aprendizado de Máquina. In.: **Data Science do Zero: noções fundamentais com Python**. 2ª edição. Rio de Janeiro: Alta Books, 2021.

# PARA SABER MAIS:

(CLIQUE NOS LINKS)



GRUS, JOEL. **Data Science do Zero: noções fundamentais com Python**. 2ª edição. Rio de Janeiro: Alta Books, 2021.

Entendendo de vez a diferença entre normalização e padronização dos dados @ Canal Ciência dos Dados

Trilha de Estudos para Cientista de Dados @ GitHub das Mulheres em Dados



# OBRIGADA PELA PARTICIPAÇÃO!

*POWERED BY MULHERES EM DADOS*

- Próxima aula em 06/07, às 19h, no Discord
- Dúvidas e sugestões no canal #python



Equipe Python:

Andressa Apio, Crislane Maria, Érika Santos e Joice Oliveira