

# Introdução Machine Learning

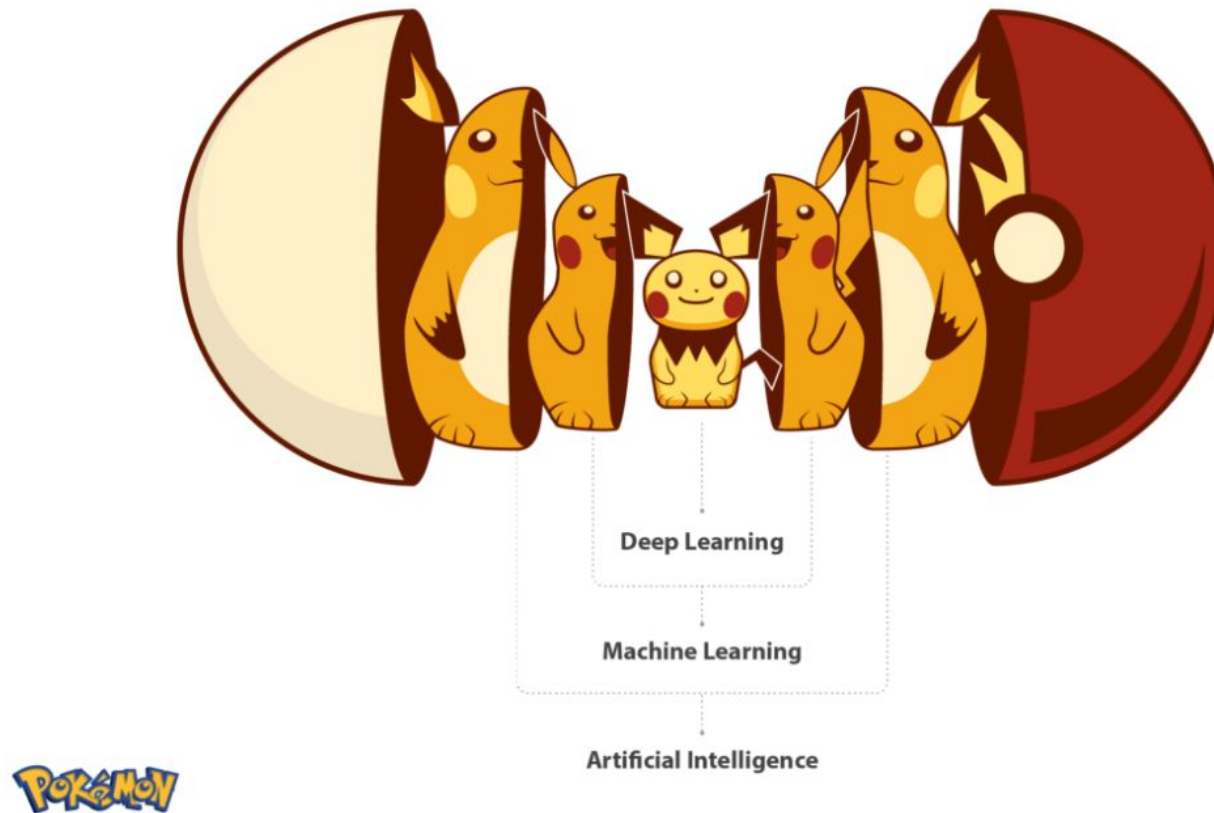
Aula 06

Professora: Dra. Rosana Rego

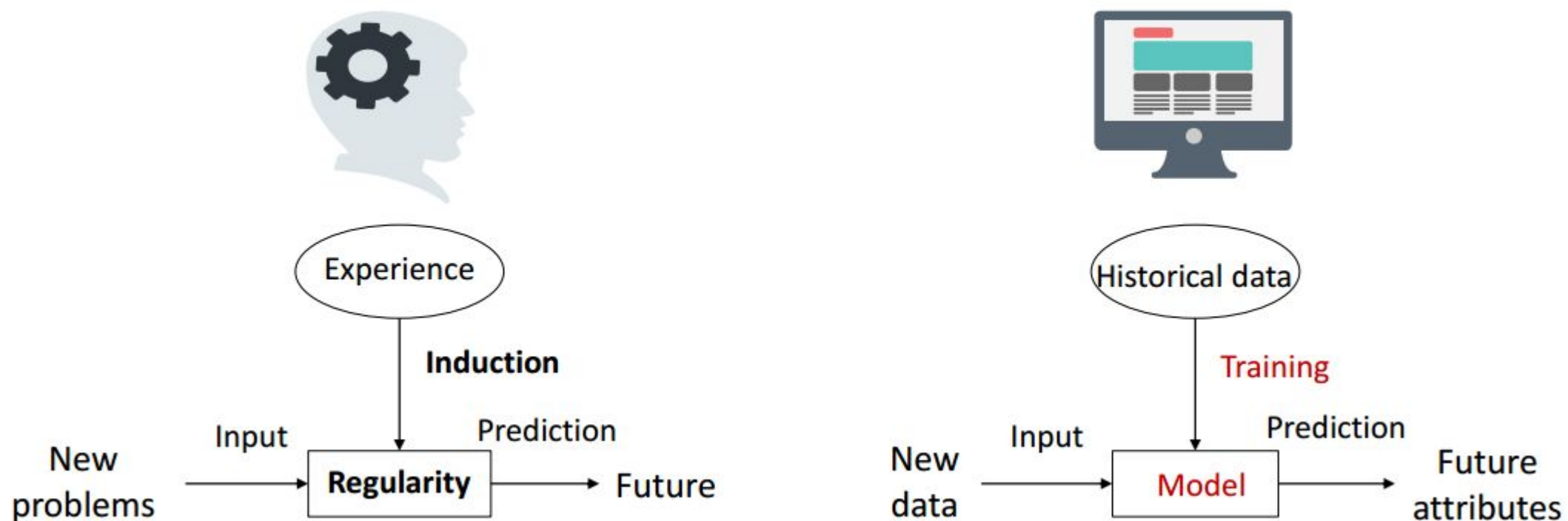
# Sumário

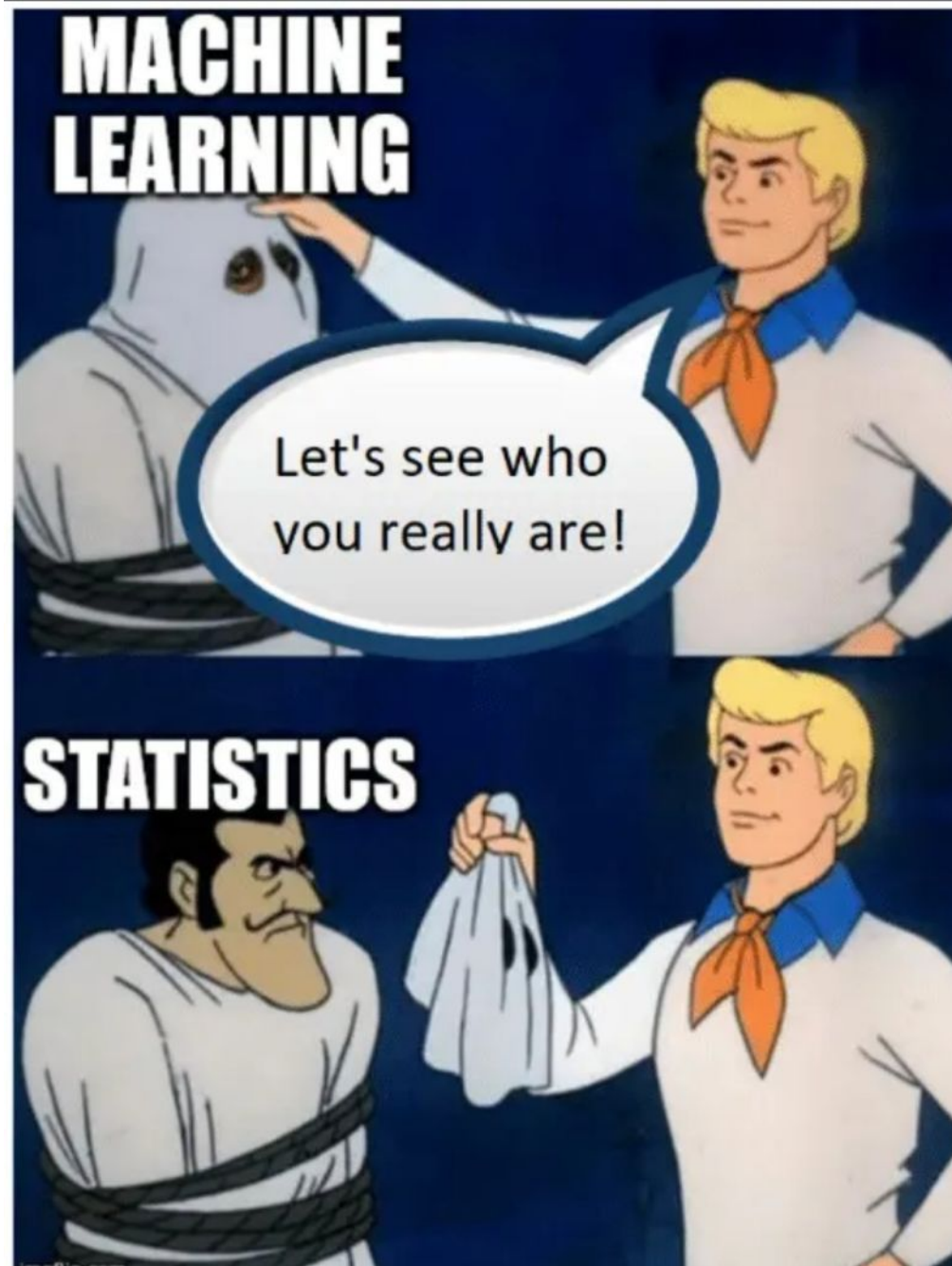
1. Introdução a machine learning
2. Workflow
3. Algoritmo Árvore de Decisão
4. Prática

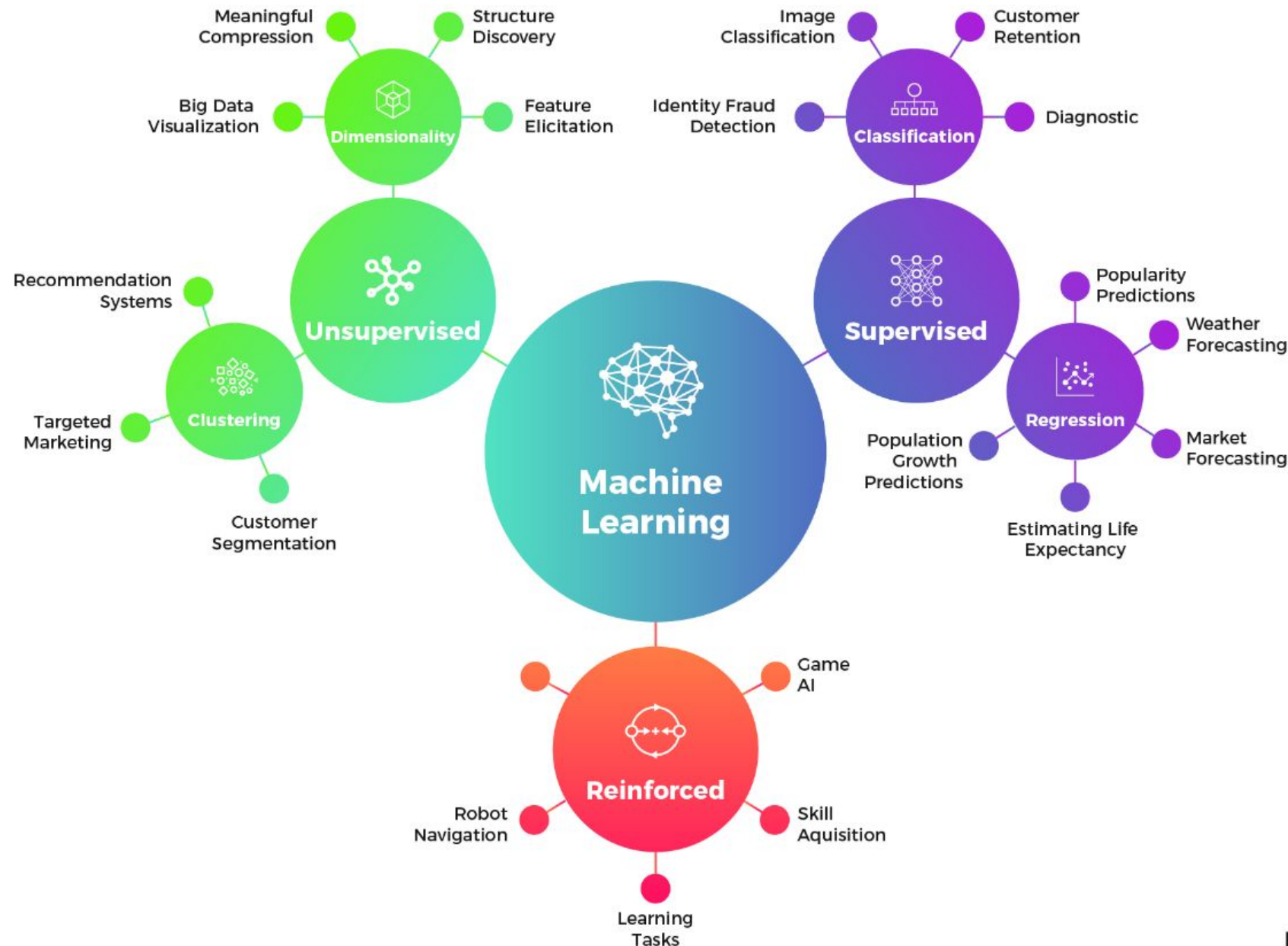
# Machine Learning (ML)



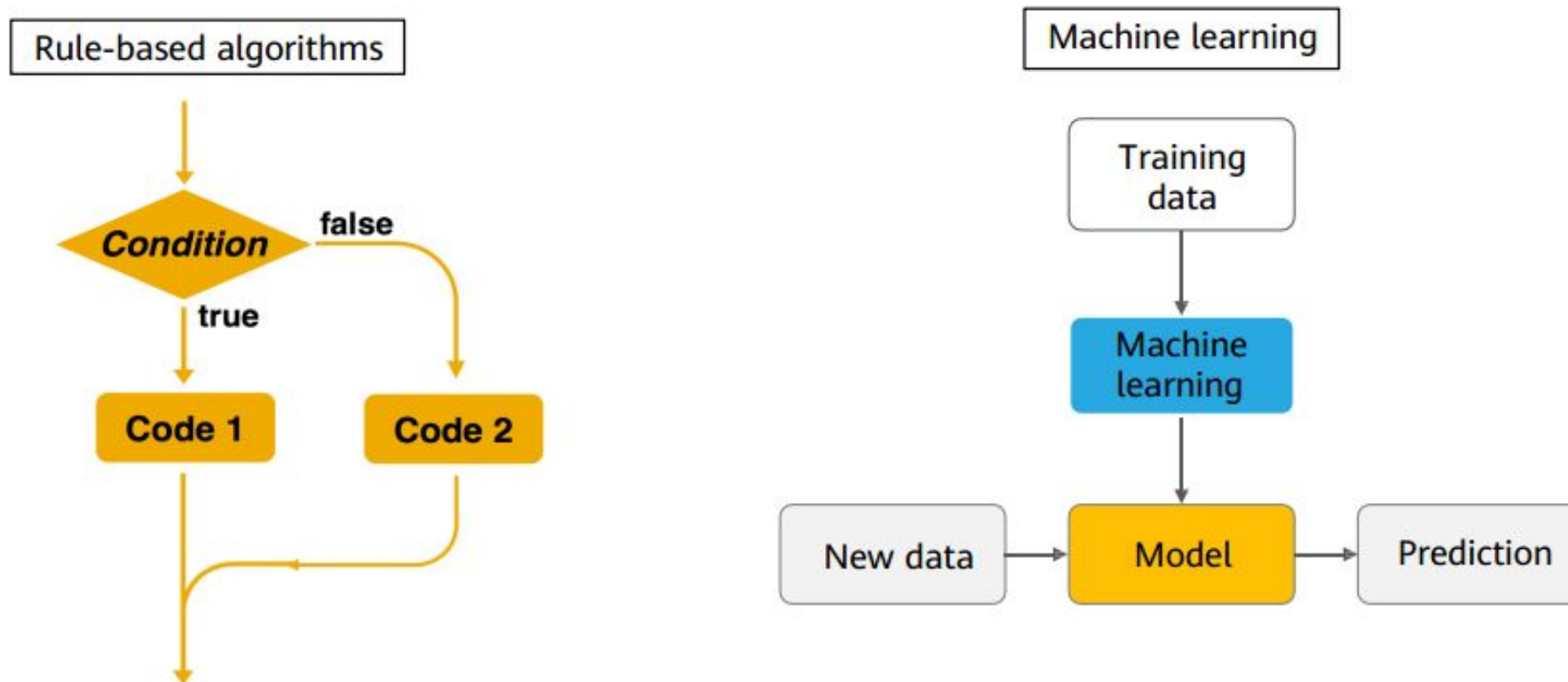
# Algoritmos de Machine Learning







# Diferenças entre algoritmos de ML e algoritmos tradicionais baseados em regras



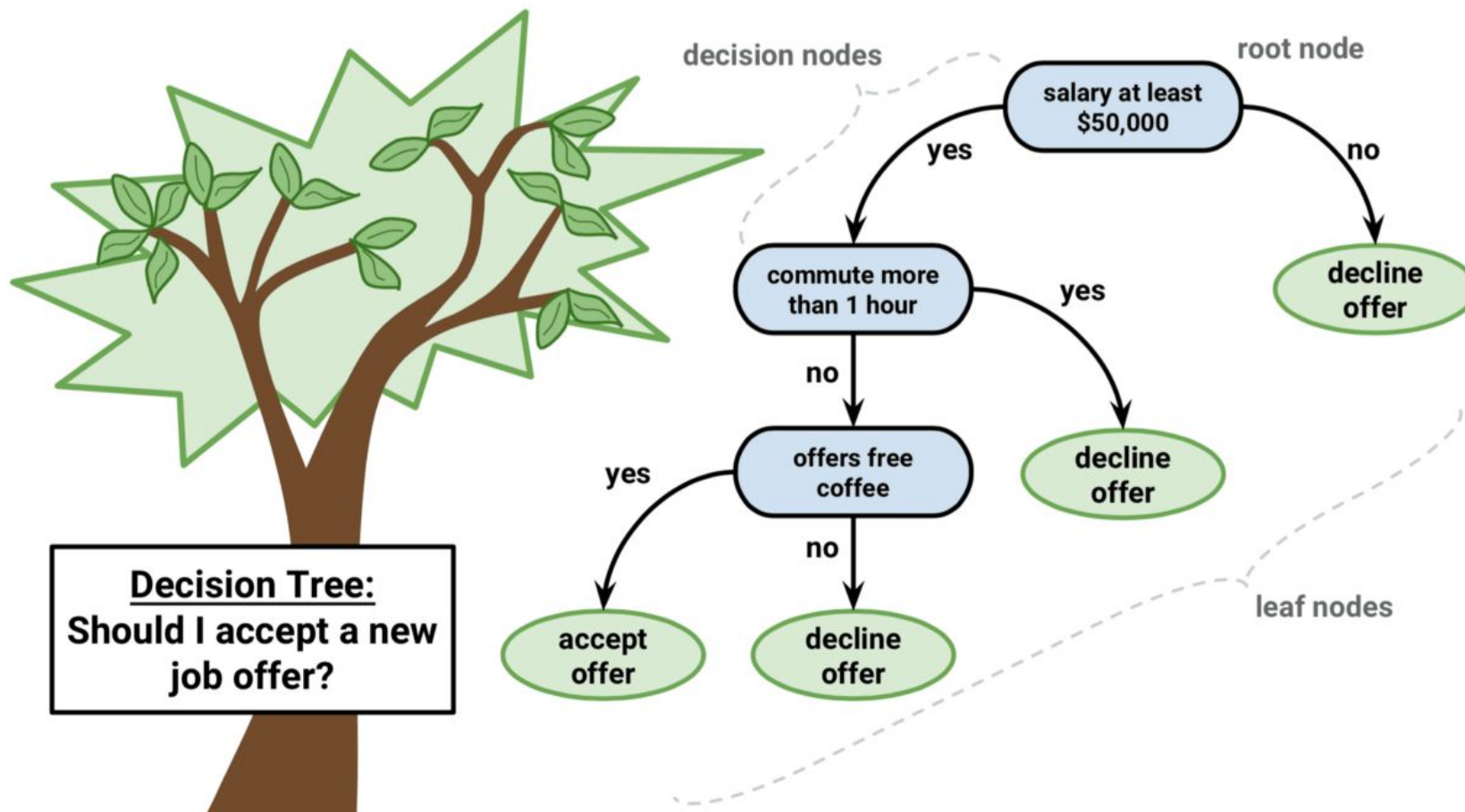
# Machine learning workflow





# EXEMPLO

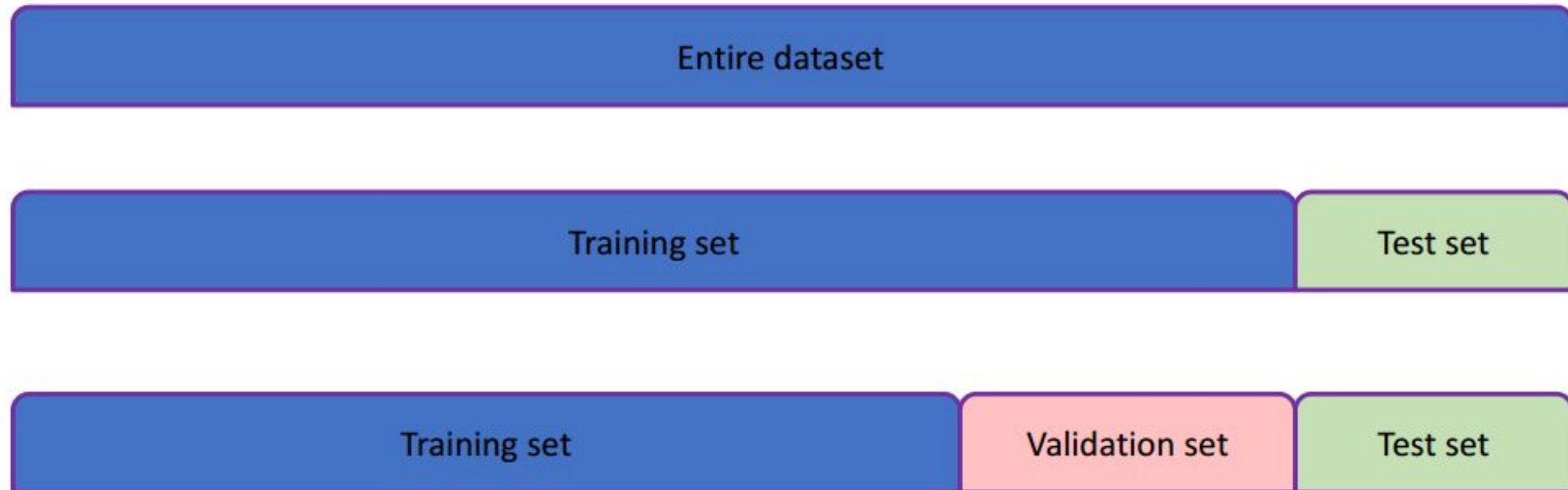
# Árvore de Decisão



# Exemplo

- Predição de hipotireoidismo com Machine Learning

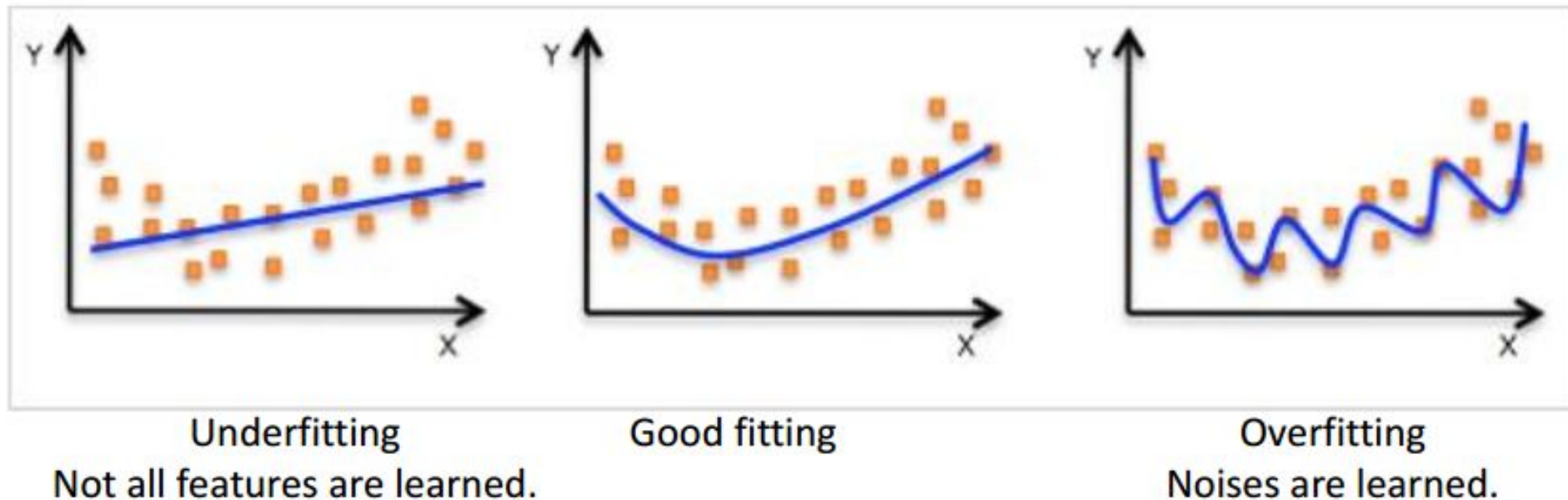
# Divisão do dataset



# Overfitting e Underfitting

# Overfitting e Underfitting

- Quando a capacidade atende à complexidade da tarefa e à quantidade de dados de treinamento fornecidos, o efeito do algoritmo geralmente é ótimo.
- Modelos com capacidade insuficiente não podem resolver tarefas complexas e pode ocorrer subajuste (*underfitting*).
- Um modelo de alta capacidade pode resolver tarefas complexas, mas pode ocorrer *overfitting* se a capacidade for maior do que a exigida por uma tarefa.



# Causa de Overfitting - Erro

Erro total da previsão final

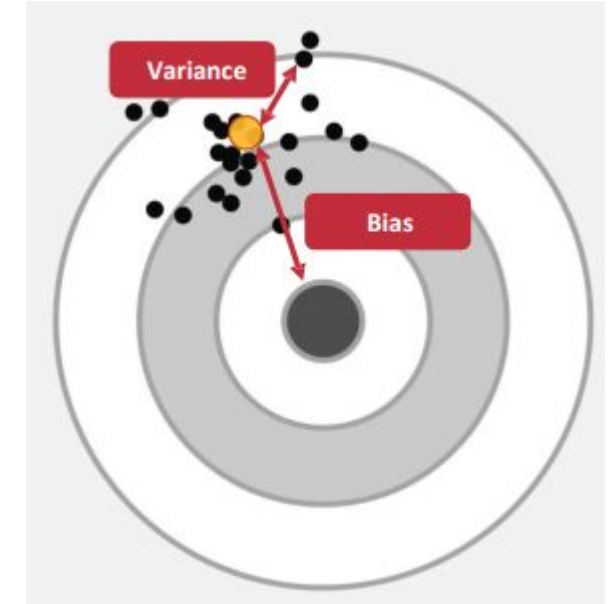
- Geralmente, o erro de previsão pode ser dividido em dois tipos:
  - Erro causado por viés/tendência
  - Erro causado pela variância

Variância

- Deslocamento do resultado da previsão do valor médio;
- Erro causado pela sensibilidade do modelo a pequenas flutuações no conjunto de treinamento.

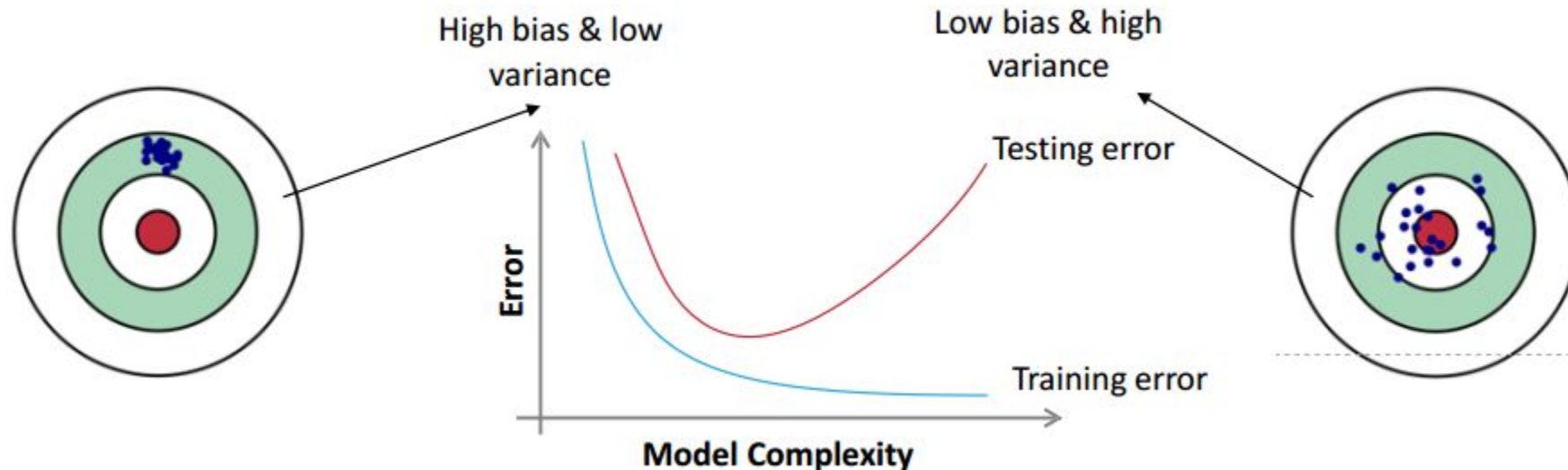
Tendência

- Diferença entre o valor de previsão esperado (ou médio) e o valor correto que estamos tentando prever.



# Variância e bias

- À medida que a complexidade do modelo aumenta, o erro de triagem diminui.
- À medida que a complexidade do modelo aumenta, o erro de teste diminui até certo ponto e depois aumenta na direção inversa, formando uma curva convexa.





# Métricas

# Métricas para avaliação do modelo: Regressão

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|$$

Quanto mais próximo o MAE estiver de 0, melhor o modelo pode ajustar os dados de treinamento.

- Mean Square Error (MSE)

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

Um erro quadrático médio de zero (0) é ideal, mas na maioria dos casos isso não é possível.

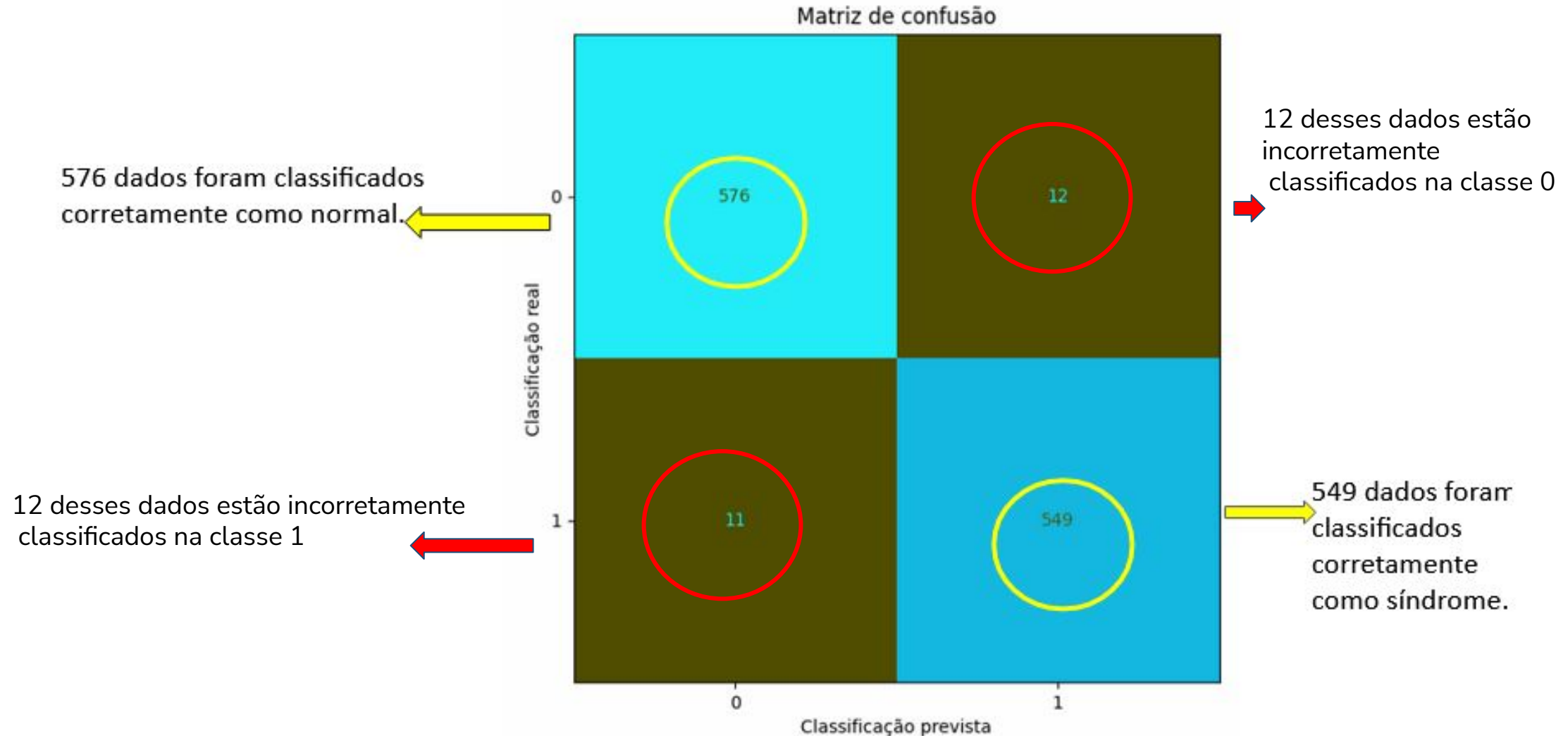
- Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

## Avaliação da performance do modelo: Classificação

- Matriz de Confusão : A matriz de confusão é uma tabela que mostra a frequência com que um classificador classifica cada classe.
  - Cada linha da matriz de confusão representa a instância de uma classe real e cada coluna representa a instância de uma classe prevista.

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

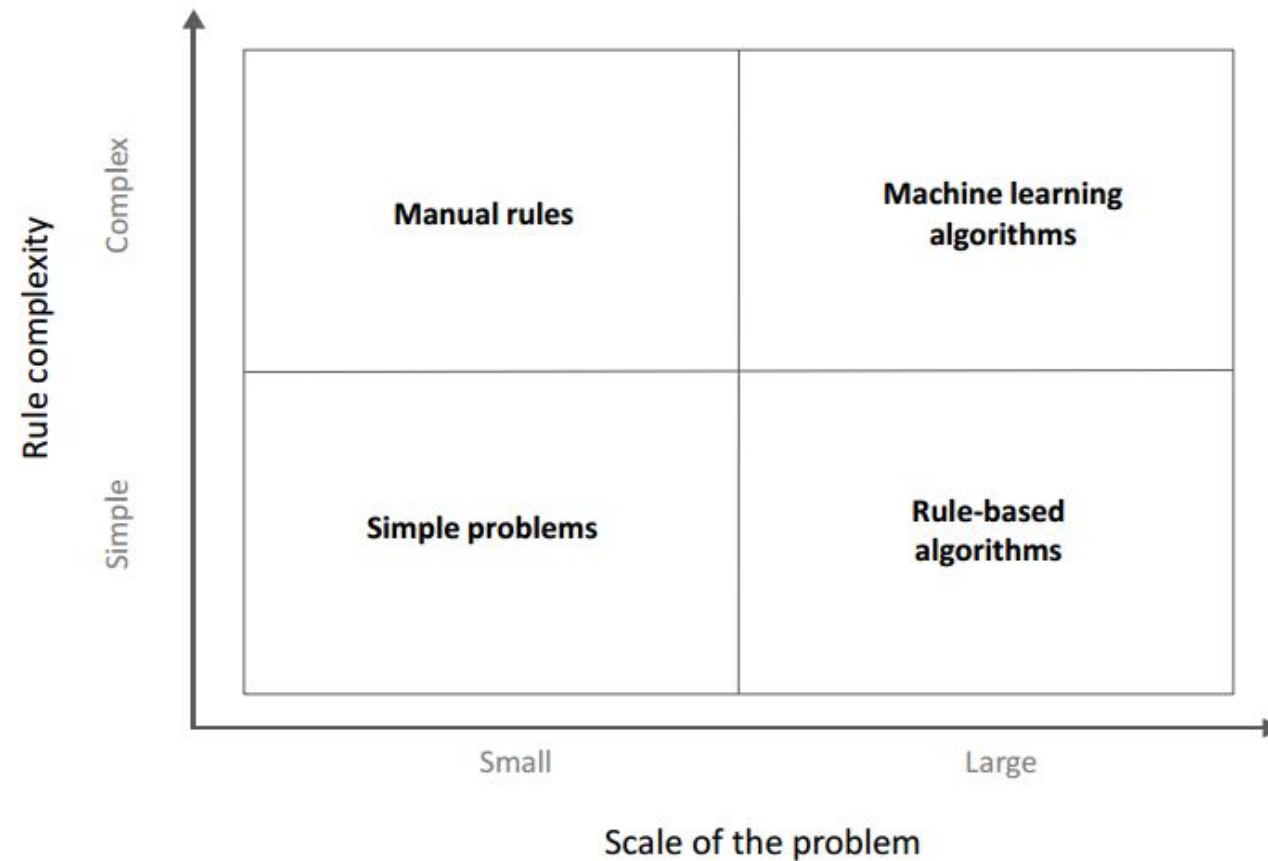


# Computational Intelligence Laboratory

## Avaliação da performance do modelo: Classificação

Measurement	Ratio
Accuracy and recognition rate	$\frac{TP + TN}{P + N}$
Error rate and misclassification rate	$\frac{FP + FN}{P + N}$
Sensitivity, true positive rate, and recall	$\frac{TP}{P}$
Specificity and true negative rate	$\frac{TN}{N}$
Precision	$\frac{TP}{TP + FP}$
$F_1$ , harmonic mean of the recall rate and precision	$\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$
$F_\beta$ , where $\beta$ is a non-negative real number	$\frac{(1 + \beta^2) \times \text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}}$

# Aplicações



# Aplicações

*Rules are complex or cannot be described, such as voice recognition.*



*Task rules change over time. For example, in the part-of-speech tagging task, new words or meanings are generated at any time.*

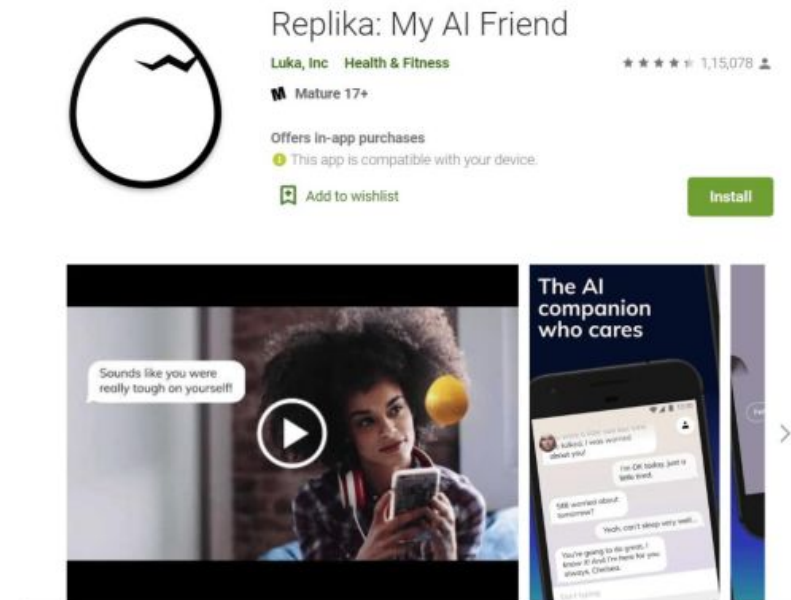


*Data distribution changes over time, requiring constant readaptation of programs, such as predicting the trend of commodity sales.*



# Aplicações - Mitigação de problemas psicológicos

Os robôs de bate-papo com IA ajudam a aliviar problemas de saúde mental, como o autismo, combinando conhecimento psicológico






# Seguro Automático de Veículos e Avaliação de Perdas

As tecnologias de IA ajudam as seguradoras a otimizar as reivindicações de seguro de veículos e concluir a avaliação de perda de seguro de veículo usando algoritmos de ML.

Vehicle Damage Assessment



```
{  
  "predictions": [  
    {  
      "label": "Dent",  
      "score": "high",  
      "xmin": 210,  
      "ymin": 139,  
      "xmax": 474,  
      "ymax": 351  
    },  
    {  
      "label": "broken",  
      "score": "high",  
      "xmin": 408,  
      "ymin": 312,  
      "xmax": 535,  
      "ymax": 399  
    },  
    {  
      "label": "Scratch",  
      "score": "medium",  
      "xmin": 639,  
      "ymin": 532,  
      "xmax": 867,  
      "ymax": 585  
    }  
  ]  
}
```

# Contato

Profa. Rosana Rego



[rosana.rego@ufersa.edu.br](mailto:rosana.rego@ufersa.edu.br)



[github.com/roscibely](https://github.com/roscibely)



[linkedin.com/in/rosanacibely/](https://www.linkedin.com/in/rosanacibely/)

# Agradecimentos

