



# Curso 2 – CD, AM e DM

## Mineração de Dados

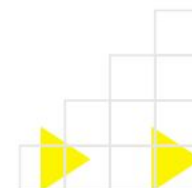
### Parte 10

#### Extração de Padrões

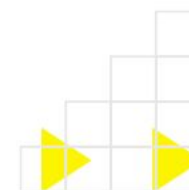
Imagens e Textos: uma breve introdução à mineração de dados não estruturados

**Prof. Ricardo M. Marcacini**

[ricardo.marcacini@icmc.usp.br](mailto:ricardo.marcacini@icmc.usp.br)



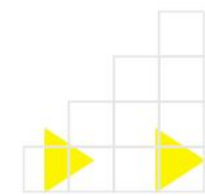
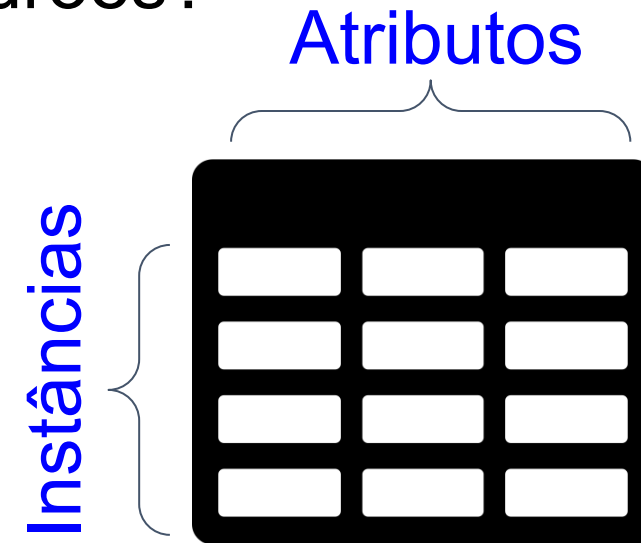
# Processo de Mineração de Dados



# Pré-processamento



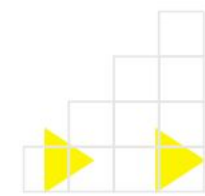
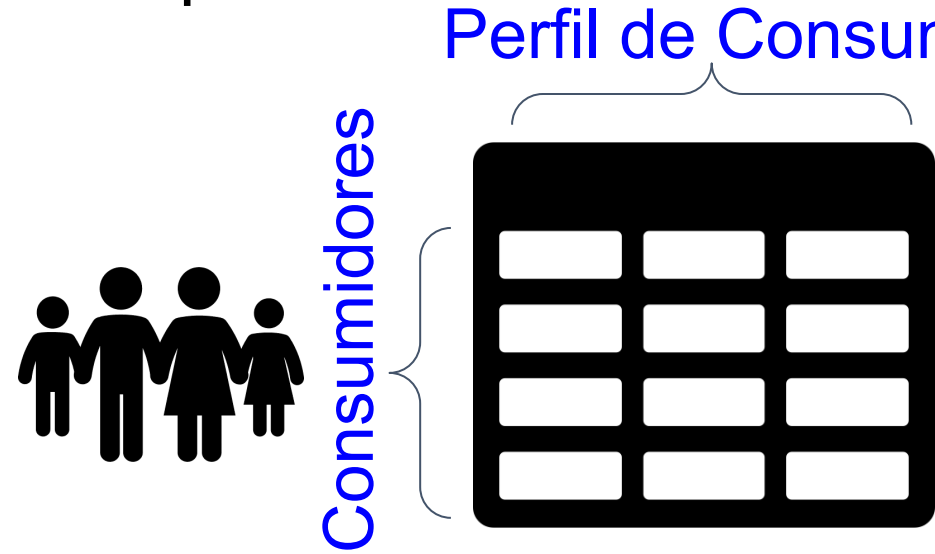
Como representar os dados para a etapa de extração de padrões?



# Pré-processamento



Como representar os dados para a etapa de extração de padrões?



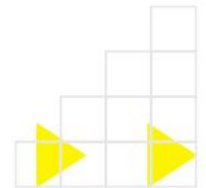
# Pré-processamento



Como representar os dados para a etapa de extração de padrões?



Textos e Imagens?

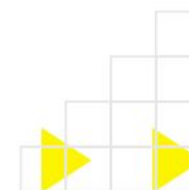




# Mineração de Dados não Estruturados



- **Atenção:** Pré-processamento de Textos e Imagens é discutido em disciplinas próprias em nosso MBA
  - Assunto complexo com diferentes técnicas para cada



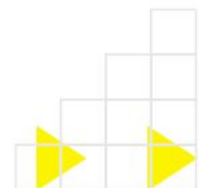
# Mineração de Dados não Estruturados



- **Atenção:** Pré-processamento de Textos e Imagens é discutido em disciplinas próprias em nosso MBA
  - Assunto complexo com diferentes técnicas para cada
  - Exemplos
    - Visão computacional
    - Processamento de Linguagem Natural

***[IMPORTANTE] Nosso objetivo:***

Por enquanto, apenas queremos testar métodos que estudamos na etapa de extração de padrões para imagens e textos!



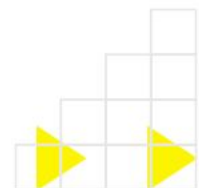
# Mineração de Dados não Estruturados



- **Atenção:** Pré-processamento de Textos e Imagens é discutido em disciplinas próprias em nosso MBA
  - Assunto complexo com diferentes técnicas para cada
  - Exemplos
    - Visão computacional
    - Processamento de Linguagem Natural

***[IMPORTANTE] Nosso objetivo:***

Vamos usar “ferramentas” prontas para pré-processamento de imagens e textos.



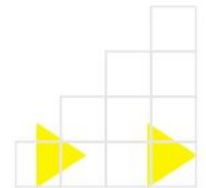
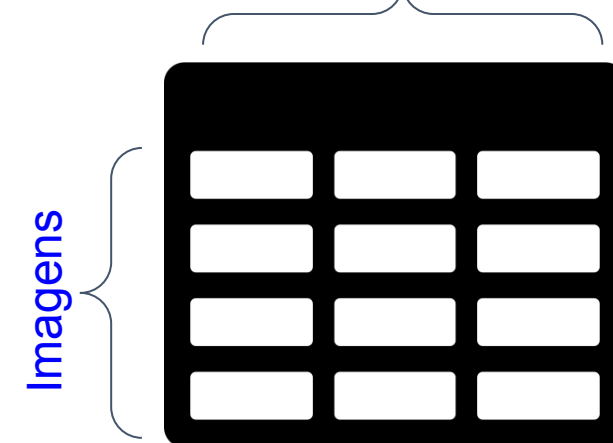


# Pré-processamento



Como representar os dados para a etapa de extração de padrões?

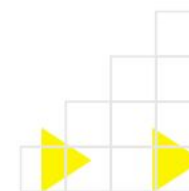
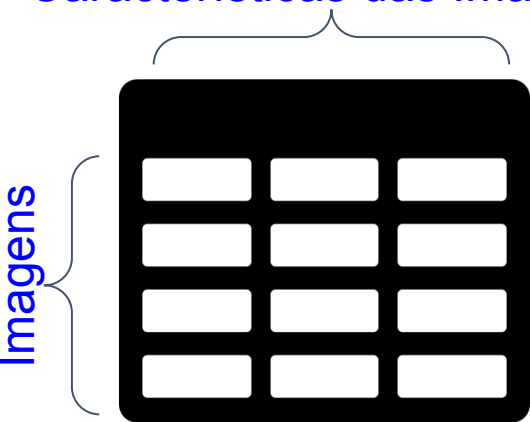
Características das Imagens



# Pré-processamento



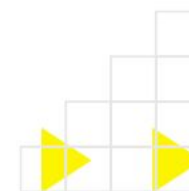
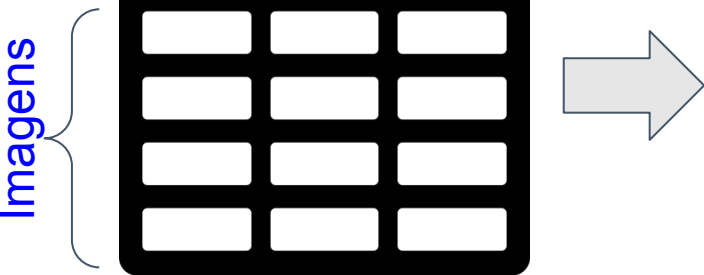
Características das Imagens



# Pré-processamento



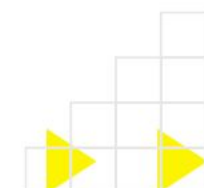
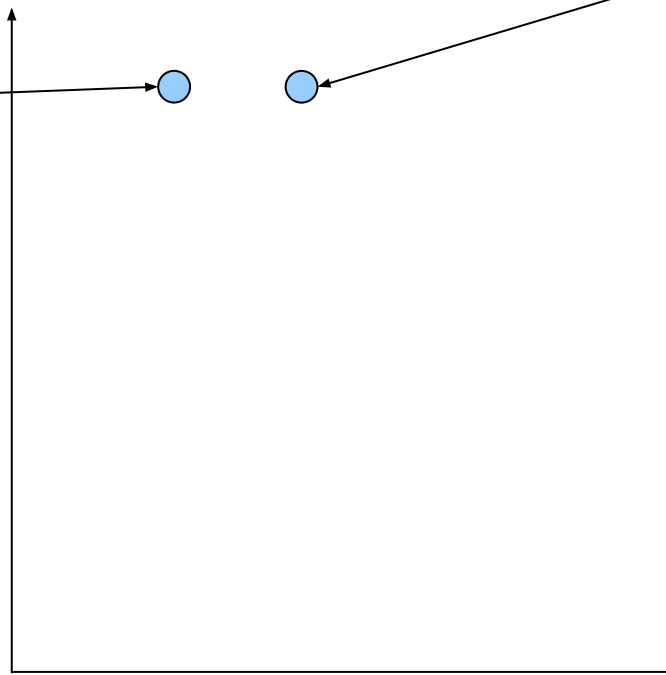
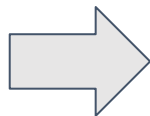
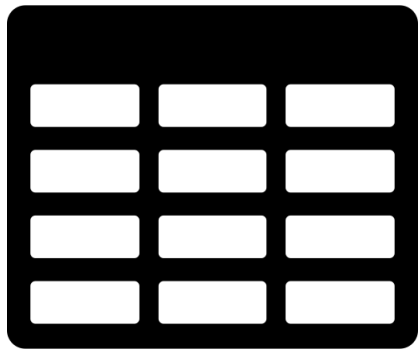
Características das Imagens



# Pré-processamento

Características das Imagens

Imagens

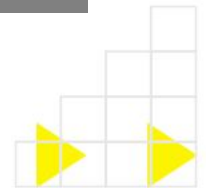
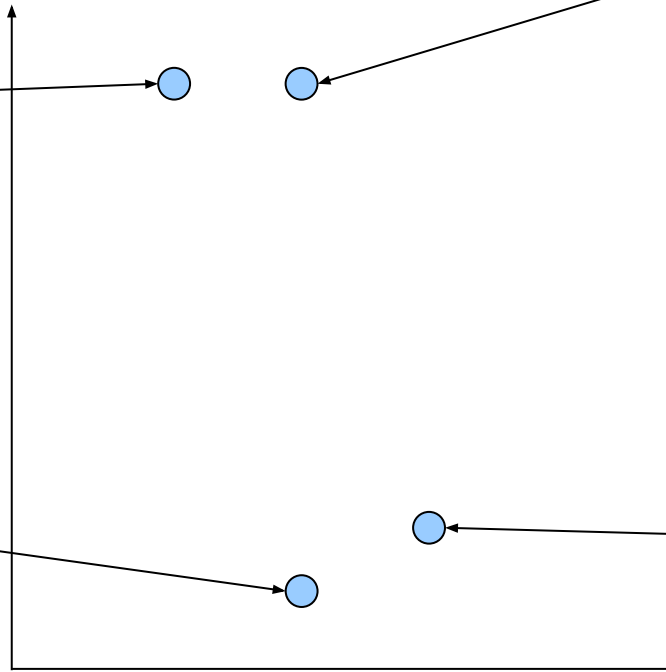
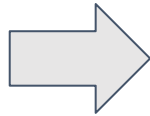
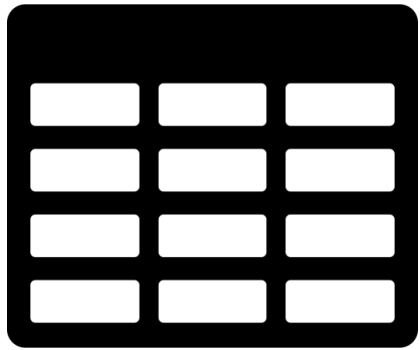


# Pré-processamento



Características das Imagens

Imagens

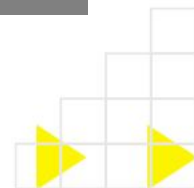
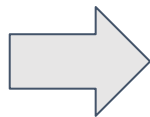
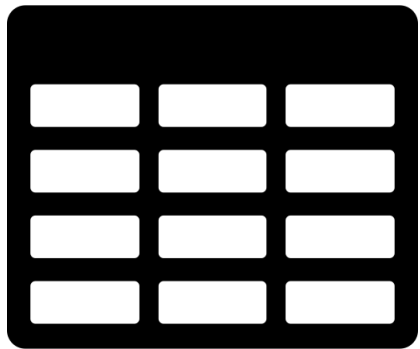


# Pré-processamento



Características das Imagens

Imagens



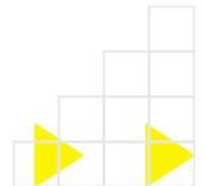
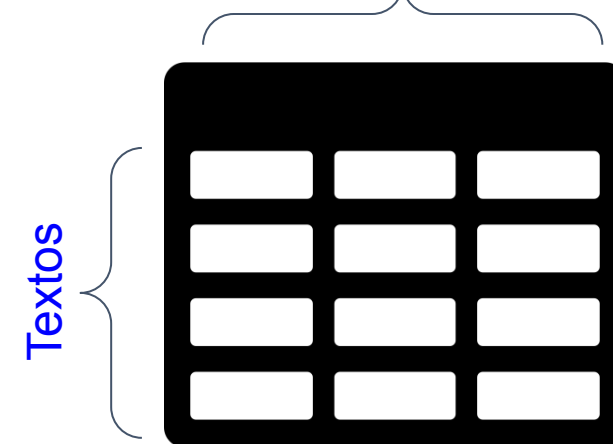


# Pré-processamento

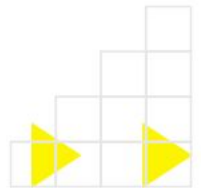
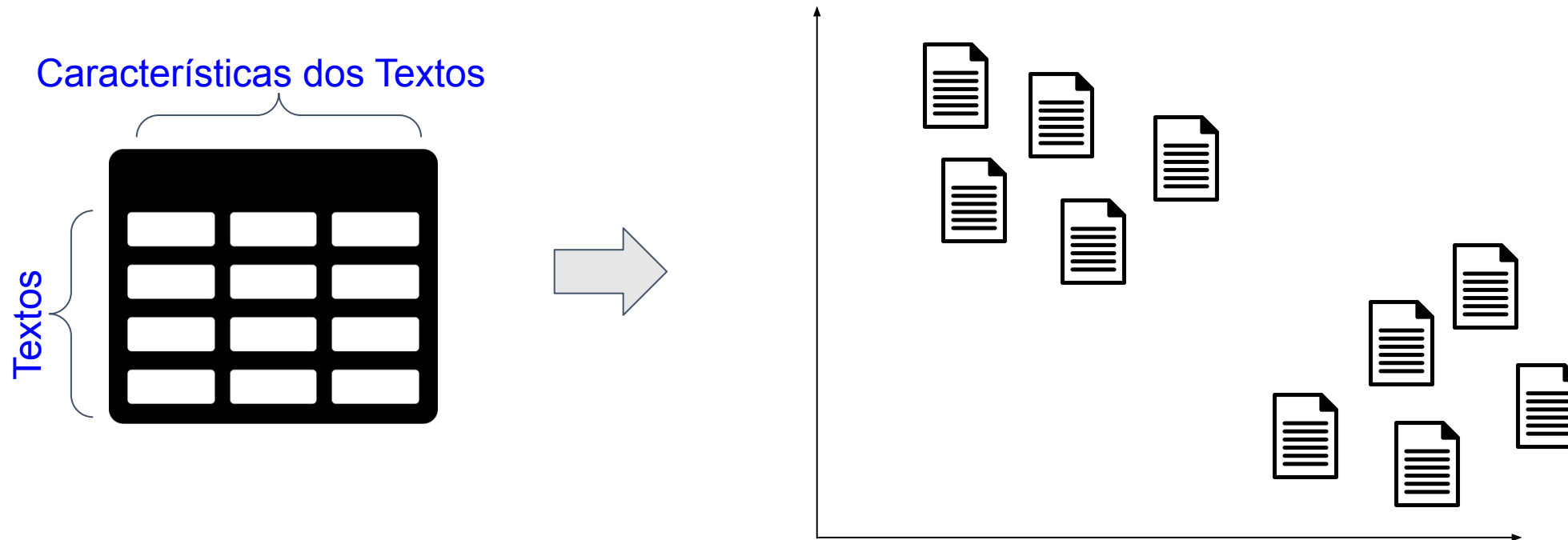


Como representar os dados para a etapa de extração de padrões?

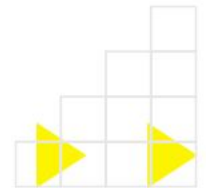
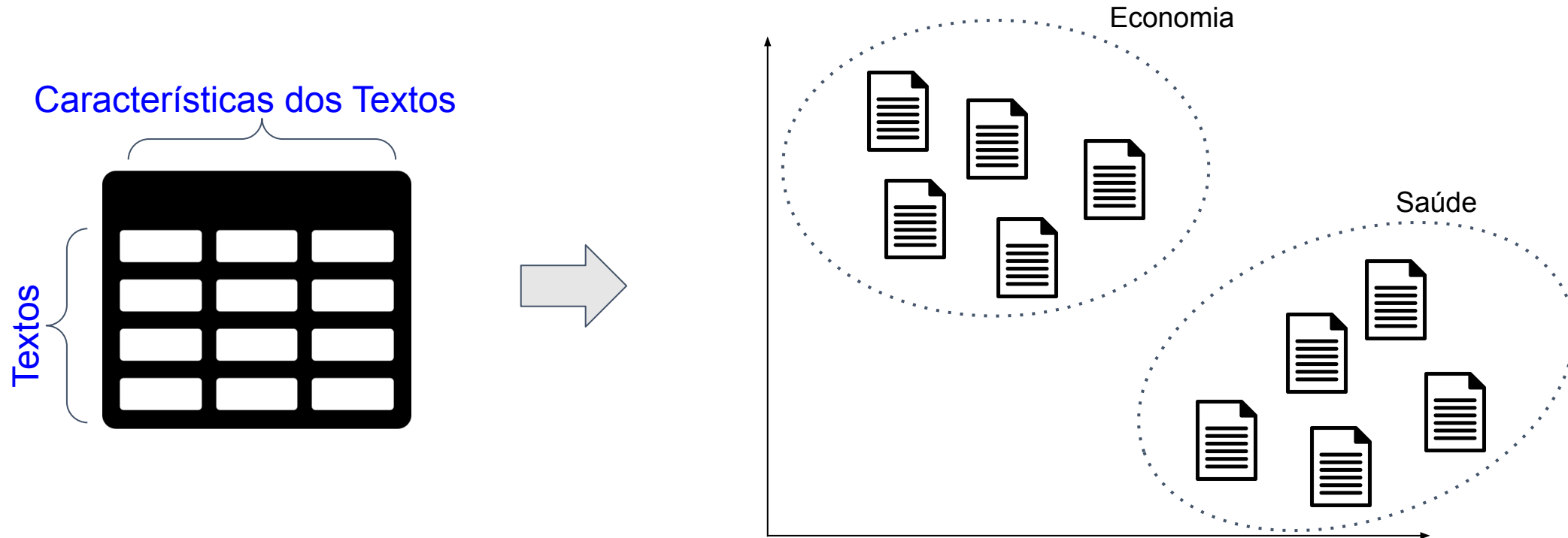
Características dos Textos



# Pré-processamento

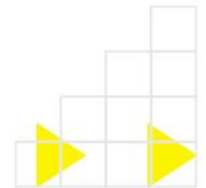
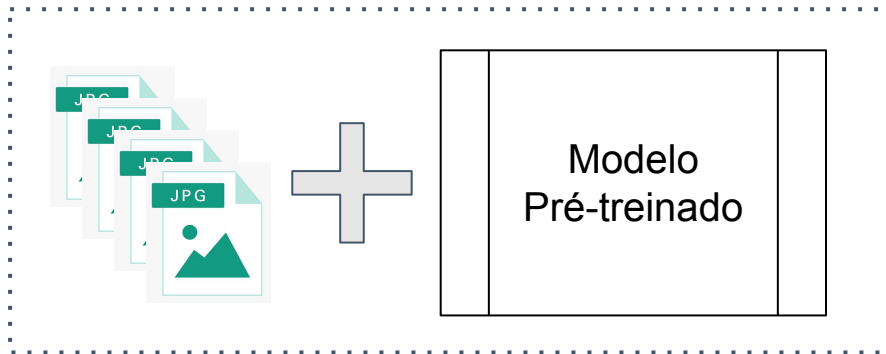


# Pré-processamento



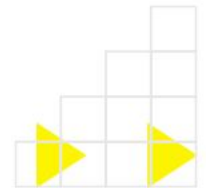
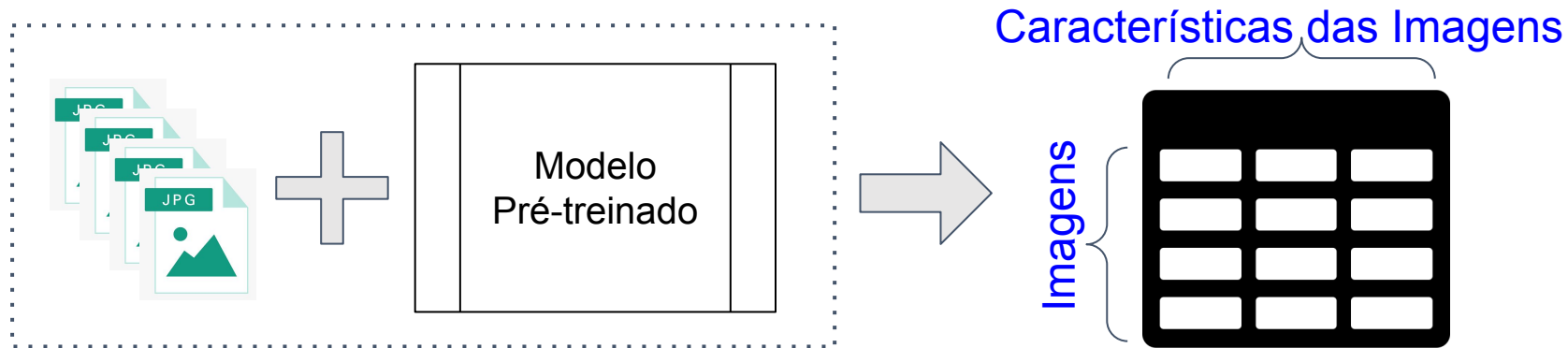
# Pré-processamento

- Em nossos estudos iniciais, vamos usar modelos pré-treinados para extração de características



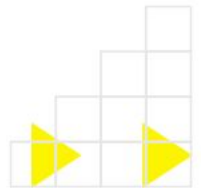
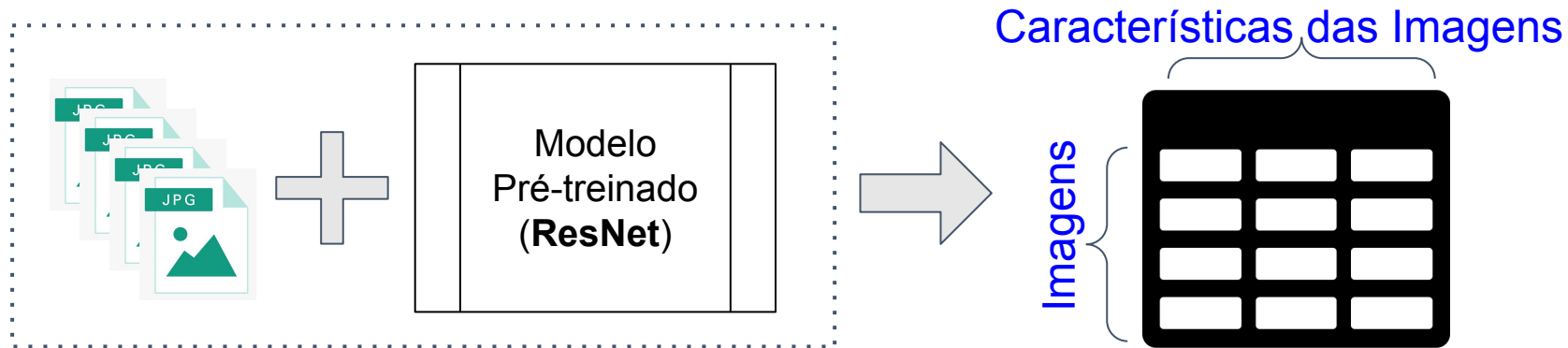
# Pré-processamento

- Em nossos estudos iniciais, vamos usar modelos pré-treinados para extração de características



# Pré-processamento

- Em nossos estudos iniciais, vamos usar modelos pré-treinados para extração de características



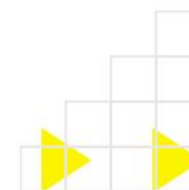
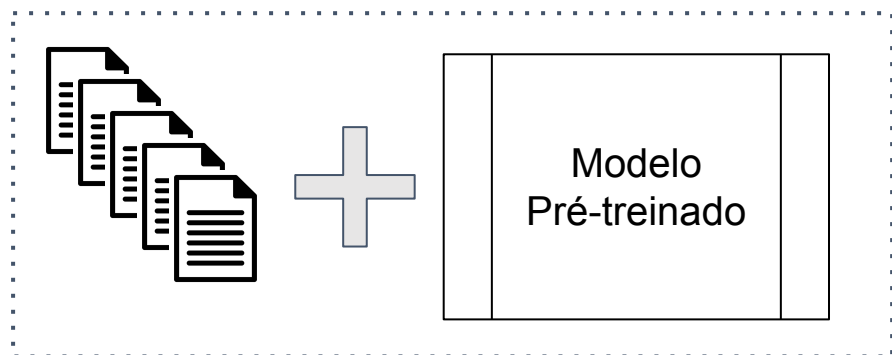
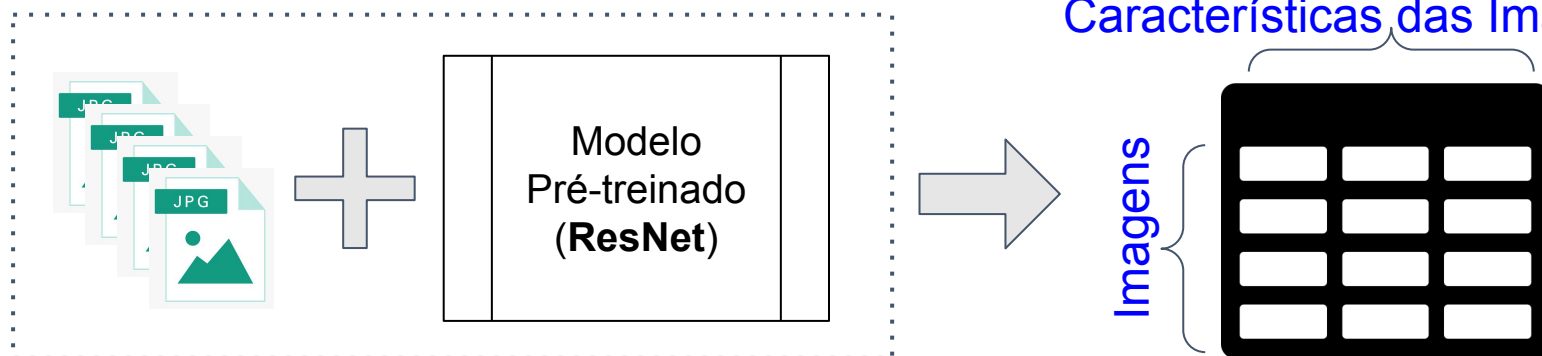


# Pré-processamento



- Em nossos estudos iniciais, vamos usar modelos pré-treinados para extração de características

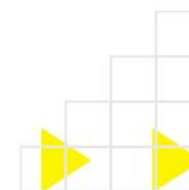
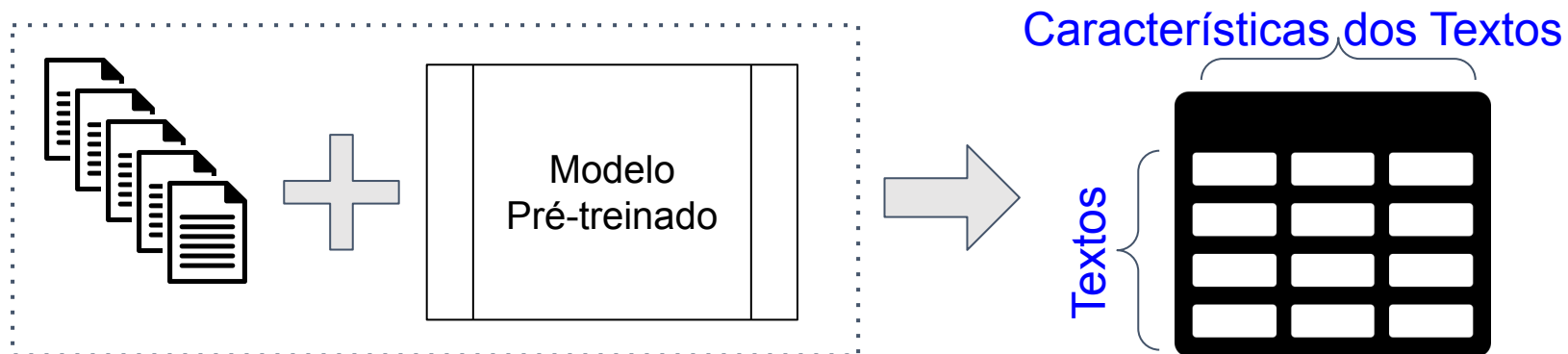
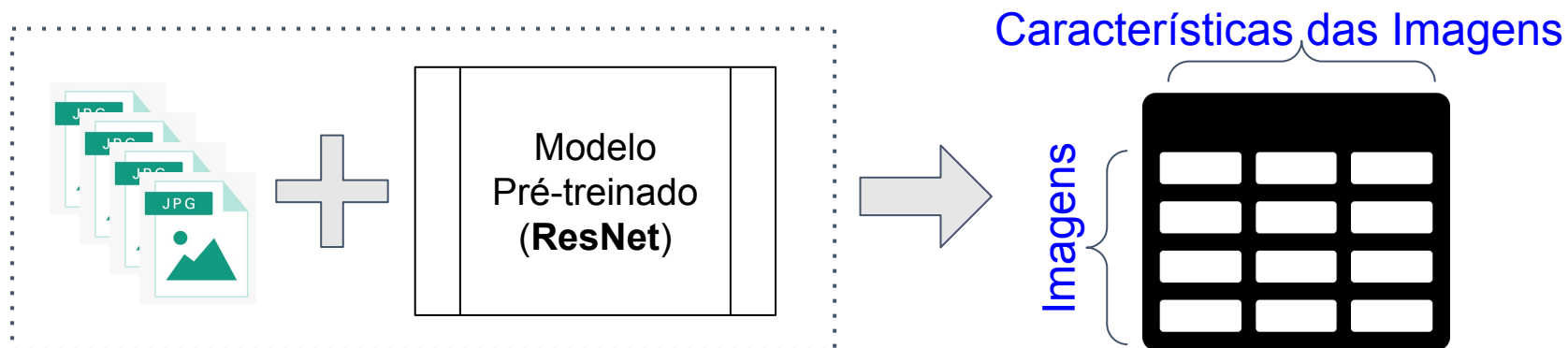
Características das Imagens



# Pré-processamento



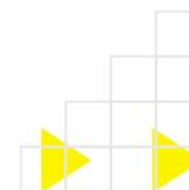
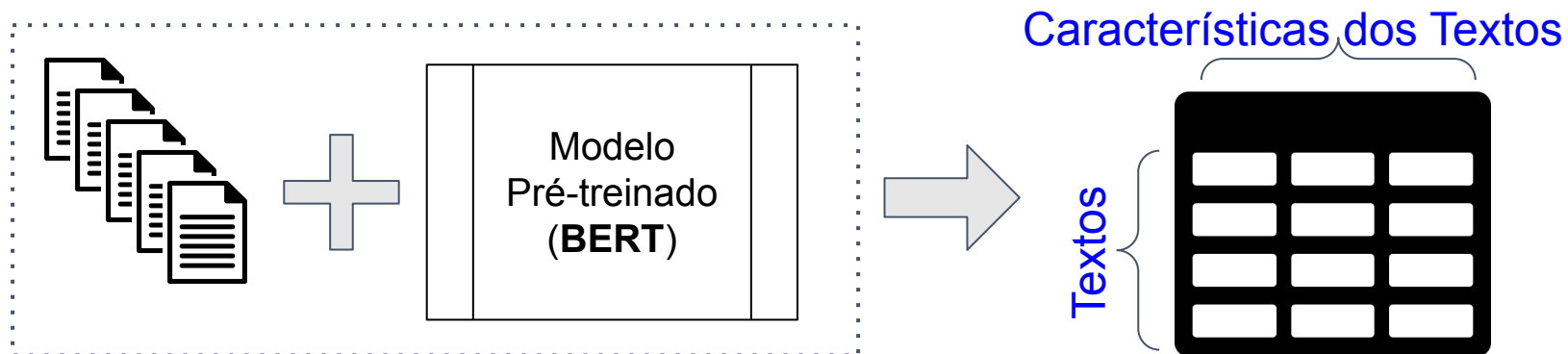
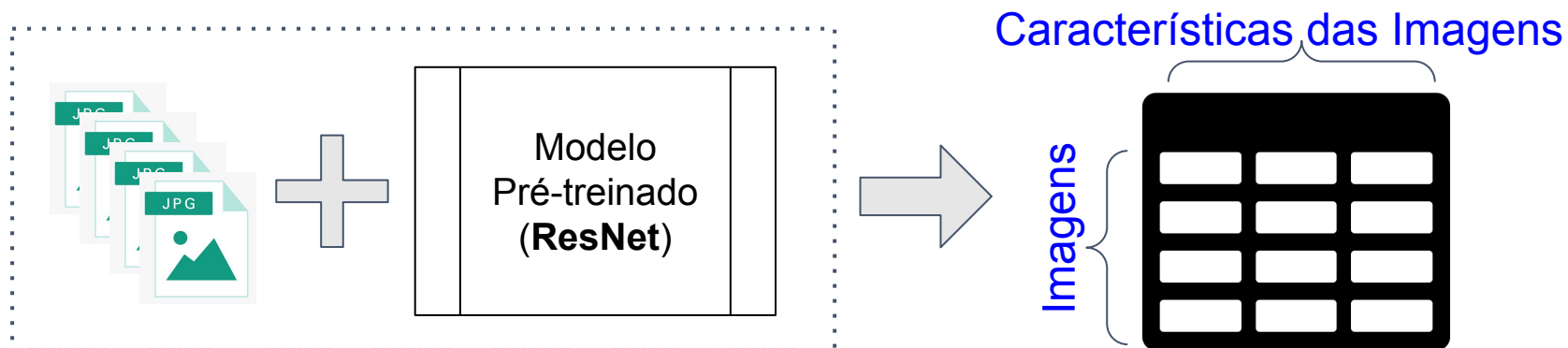
- Em nossos estudos iniciais, vamos usar modelos pré-treinados para extração de características



# Pré-processamento

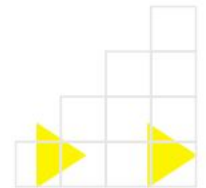
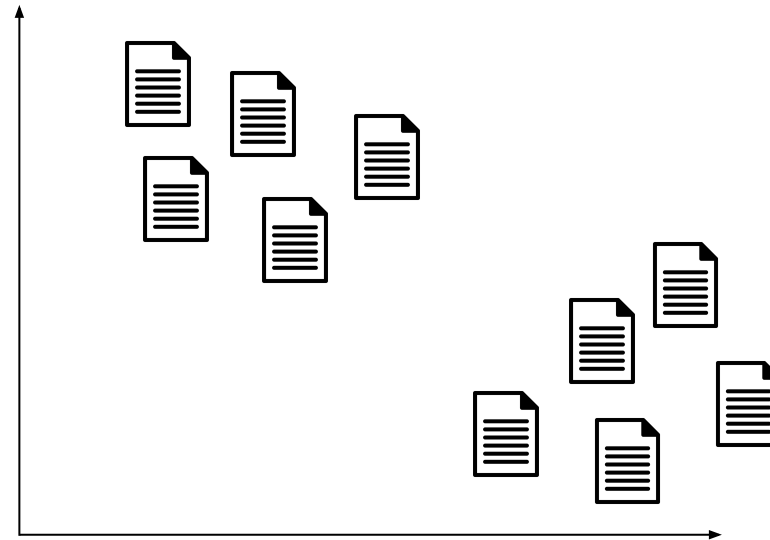
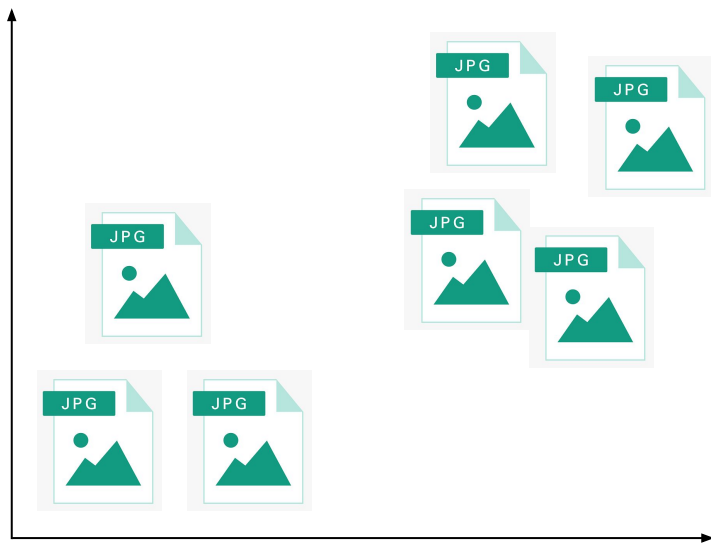


- Em nossos estudos iniciais, vamos usar modelos pré-treinados para extração de características



# Pré-processamento

- Em nossos estudos iniciais, vamos usar modelos pré-treinados para extração de características



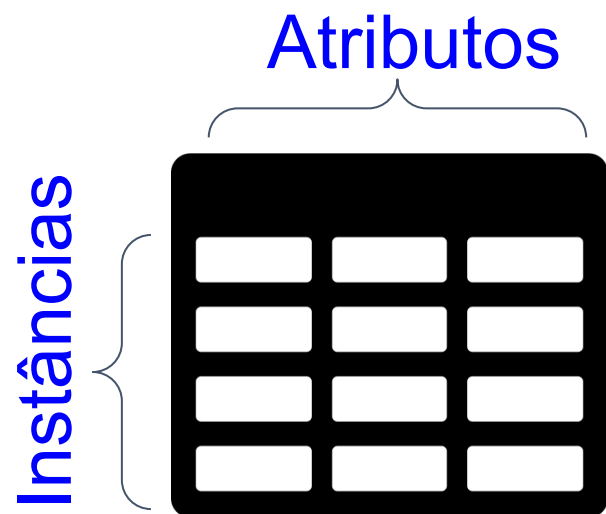
# Pré-processamento



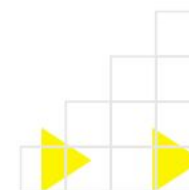
- Imagens e textos pré-processados também são representados no modelo espaço-vetorial!

(Aula: Parte 2) cada objeto será denotado por um vetor:

$$\mathbf{x}_i = [x_{i1} \cdots x_{id}]$$



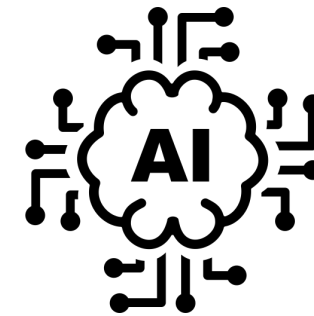
$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix}$$



# Extração de Padrões

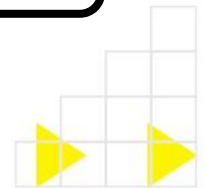


## Métodos de Aprendizado de Máquina



Tarefas Preditivas

Tarefas Descritivas







# Extração de Padrões



## Tarefas Preditivas

*Classificação de Textos usando  $k$ -NN*

## Tarefas Descritivas

*Agrupamento de Imagens usando  $k$ -Means*

