



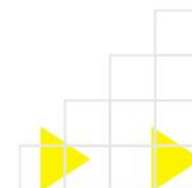
# Curso 2 – CD, AM e DM

## Mineração de Dados

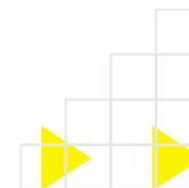
### Parte 1

### Etapas da Mineração de Dados

**Prof. Ricardo M. Marcacini**  
[ricardo.marcacini@icmc.usp.br](mailto:ricardo.marcacini@icmc.usp.br)



# Processo de Mineração de Dados

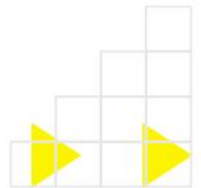


# Processo de Mineração de Dados



## Objetivo

Extração de conhecimento útil e interessante, a partir de dados, para a utilização em um processo de tomada de decisão.

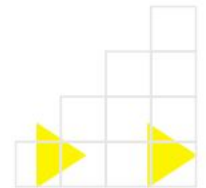


# Processo de Mineração de Dados



## Interdisciplinar

- Banco de Dados
- Aprendizado de Máquina
- Estatística
- Visualização Computacional

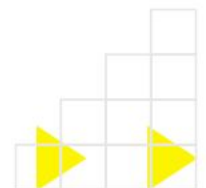


# Processo de Mineração de Dados



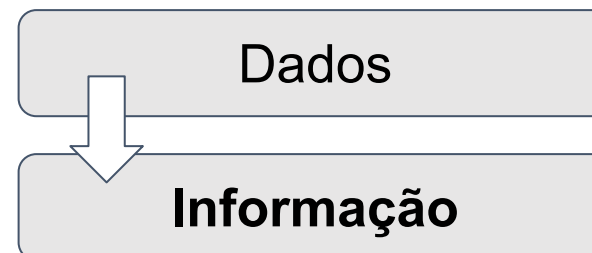
Dados

9	88	21





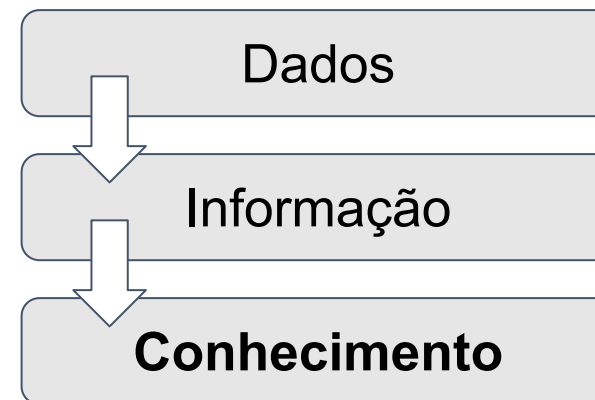
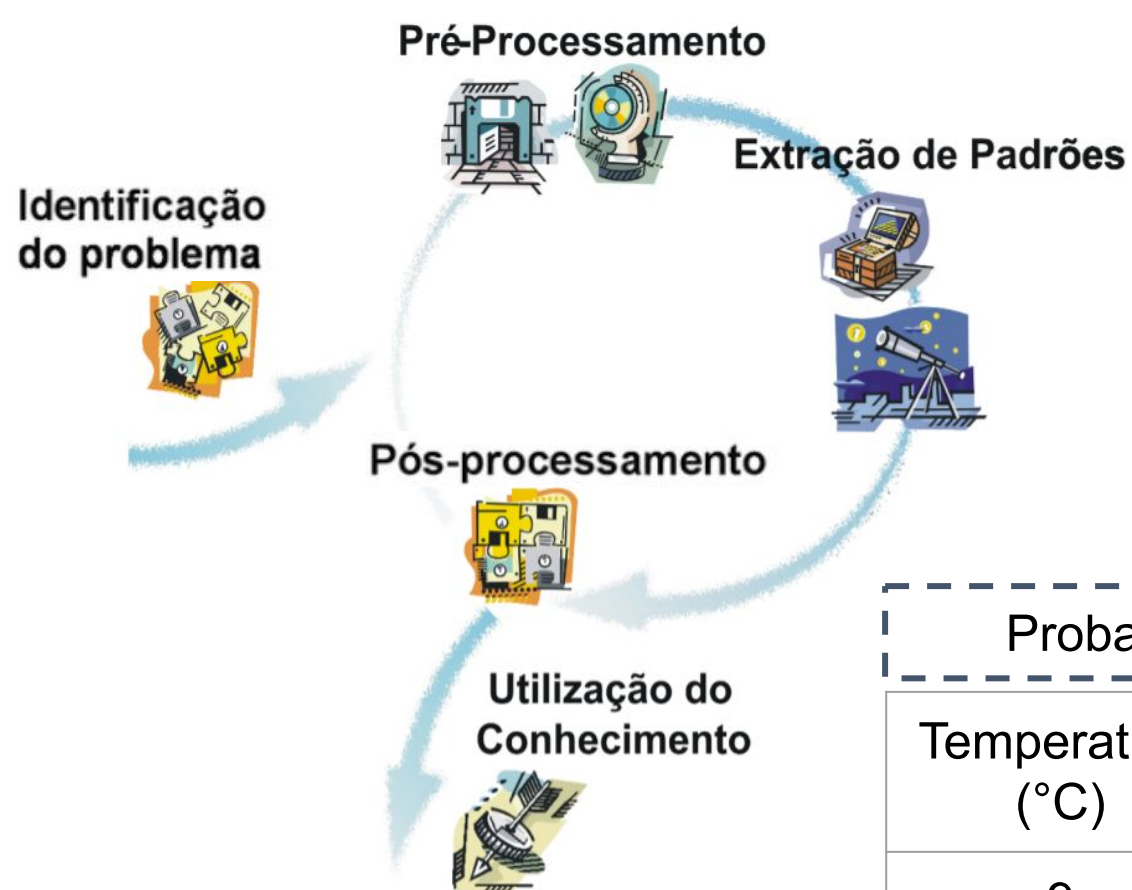
# Processo de Mineração de Dados



Temperatura (°C)	Umidade (%)	Vento (km/h)
9	88	21



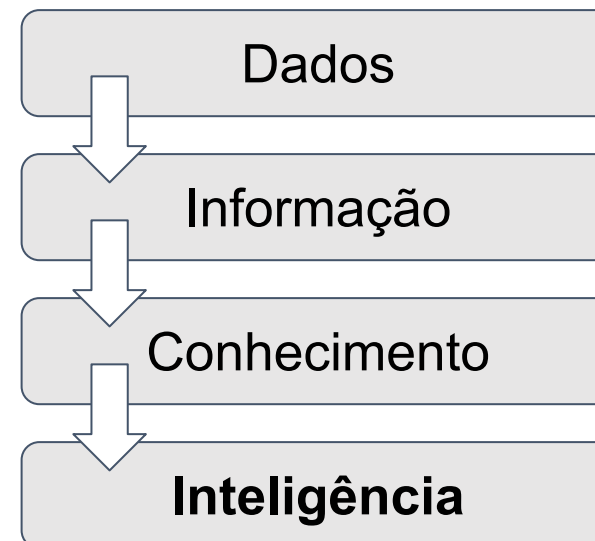
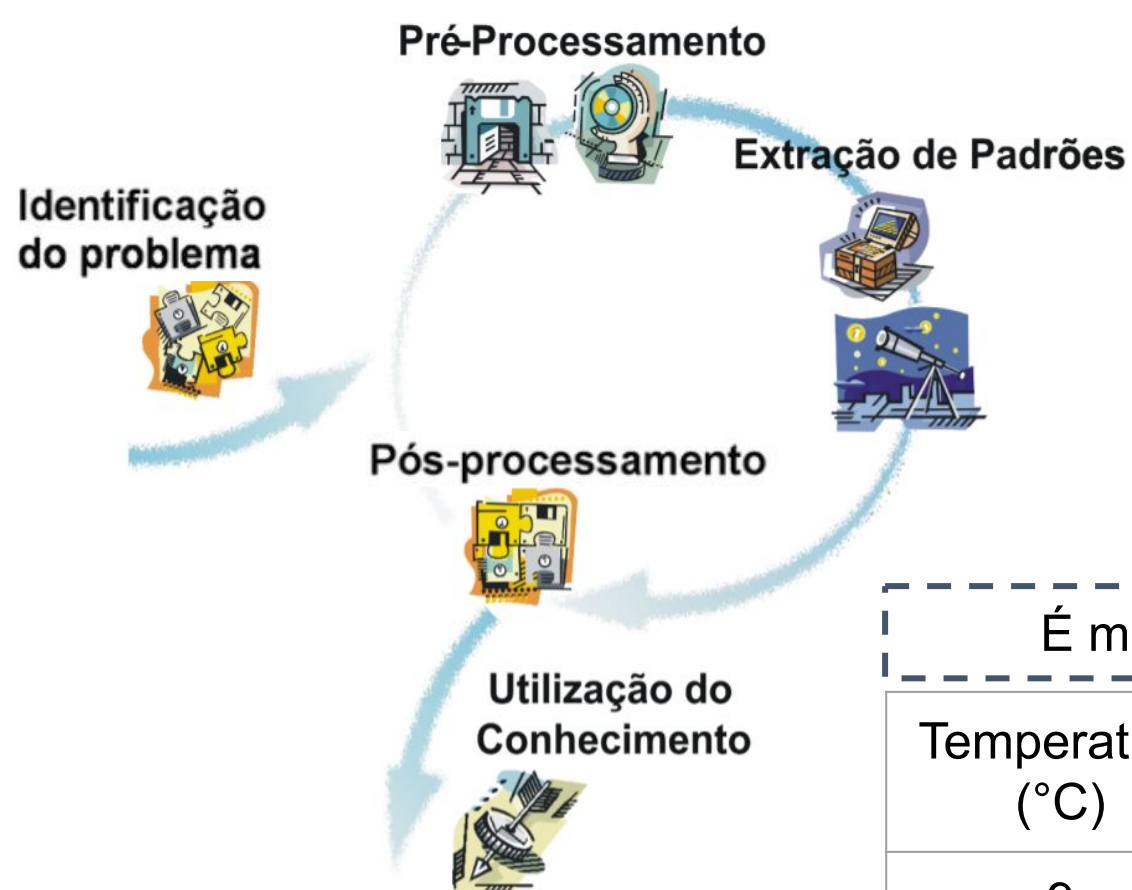
# Processo de Mineração de Dados



Probabilidade de chuva forte = 80%		
Temperatura (°C)	Umidade (%)	Vento (km/h)
9	88	21



# Processo de Mineração de Dados



É melhor levar um guarda-chuva ☺		
Temperatura (°C)	Umidade (%)	Vento (km/h)
9	88	21



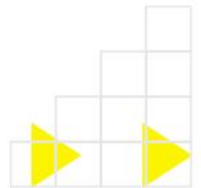


# Processo de Mineração de Dados



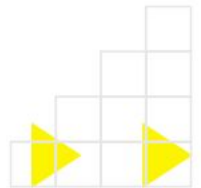
## Objetivo

Extração de conhecimento útil e interessante, a partir de dados, para a utilização em um processo de tomada de decisão.



# Identificação do Problema

Qual é o objetivo?

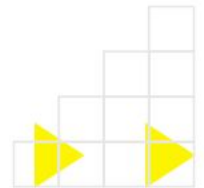
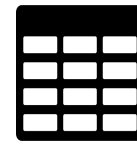


# Identificação do Problema



Qual é o objetivo?

Quais os dados disponíveis?



# Identificação do Problema

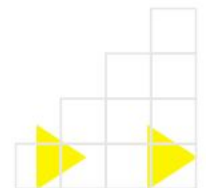


Qual é o objetivo?

Quais os dados disponíveis?



Quais são as metas?



# Identificação do Problema

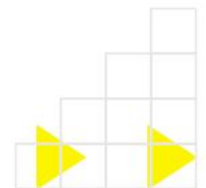
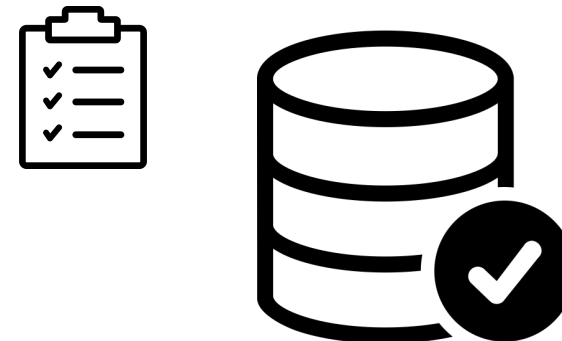


Qual é o objetivo?

Quais os dados disponíveis?



Quais são as metas?

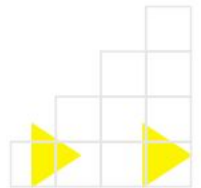




# Pré-processamento



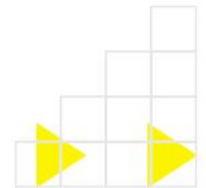
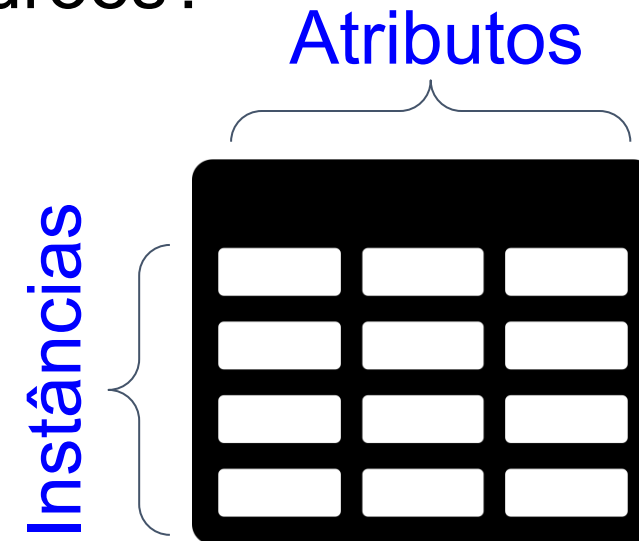
Como representar os dados para a etapa de extração de padrões?



# Pré-processamento



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

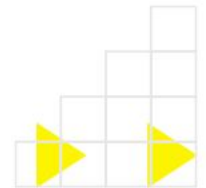
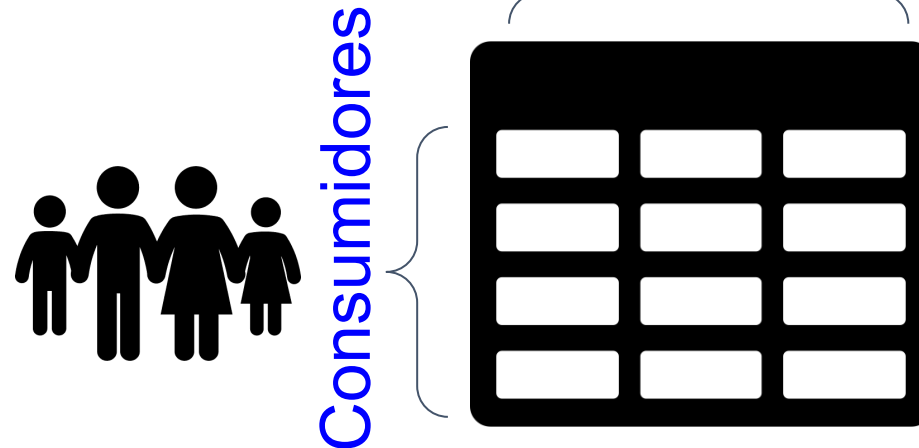


# Pré-processamento



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

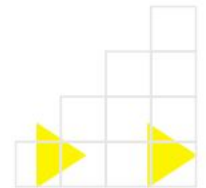
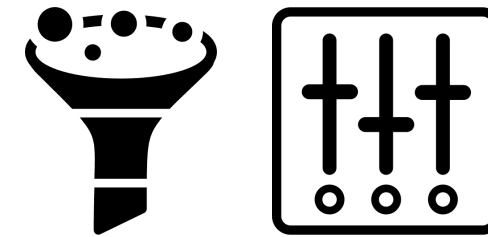
Perfil de Consumo



# Pré-processamento



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



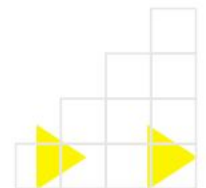
# Pré-processamento



Filtro e Limpeza

Valores Ausentes

Valores Inconsistentes





# Pré-processamento



Filtro e Limpeza

Valores Ausentes

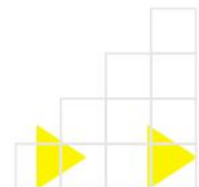
Valores Inconsistentes

Transformação de Dados

Conversão de Valores

Normalização

Padronização



# Pré-processamento



Filtro e Limpeza

Valores Ausentes

Valores Inconsistentes

Transformação de Dados

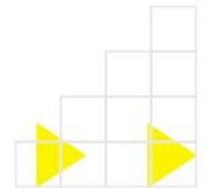
Conversão de Valores

Normalização

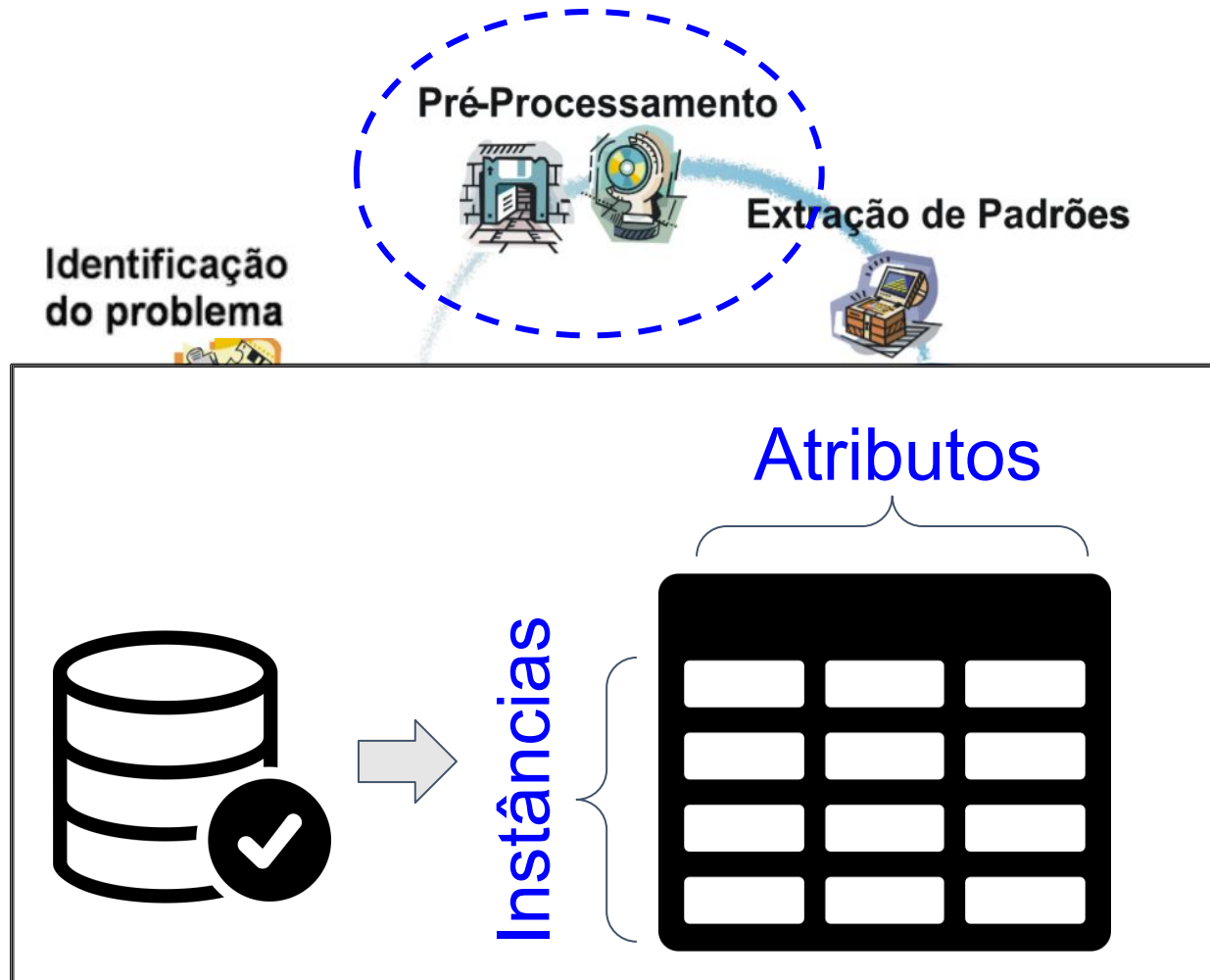
Padronização

Redução da Dimensionalidade

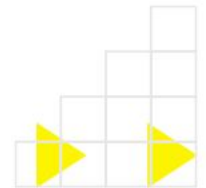
Balanceamento de Dados



# Pré-processamento



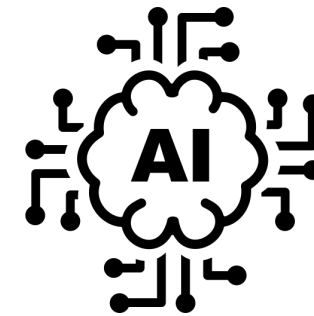
- Filtro e Limpeza
- Valores Ausentes
- Valores Inconsistentes
- Transformação de Dados
- Conversão de Valores
- Normalização
- Padronização
- Redução da Dimensionalidade
- Balanceamento de Dados



# Extração de Padrões

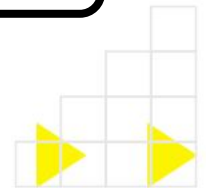


## Métodos de Aprendizado de Máquina



Tarefas Preditivas

Tarefas Descritivas



# Extração de Padrões



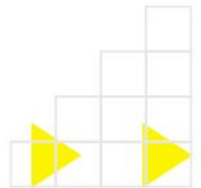
## Aprendizado Supervisionado

Tarefas Preditivas

Exemplos já estudados:

Árvores de Decisão

Regressão





# Extração de Padrões

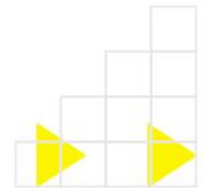


Aprendizado não  
Supervisionado

Tarefas Descritivas

Agrupamento de Dados

Regras de Associação

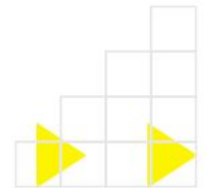


# Pós-Processamento



As metas foram atingidas?

Quais os critérios de avaliação?



# Pós-Processamento



As metas foram atingidas?

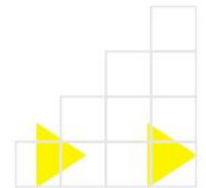
Quais os critérios de avaliação?

Exemplos:

Acurácia

Precisão e Revocação

Validação Cruzada

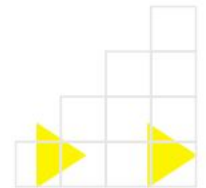


# Utilização do Conhecimento

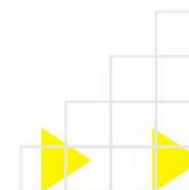


## Exemplos de aplicação para o uso do conhecimento

- Marketing / e-commerce
- Medicina
- Financeira
- Mídia / Redes Sociais



# Mineração de Dados





# Mineração de Dados



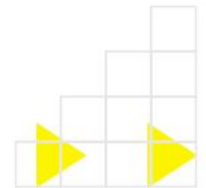
## Tarefas Descritivas

Agrupamento de Dados

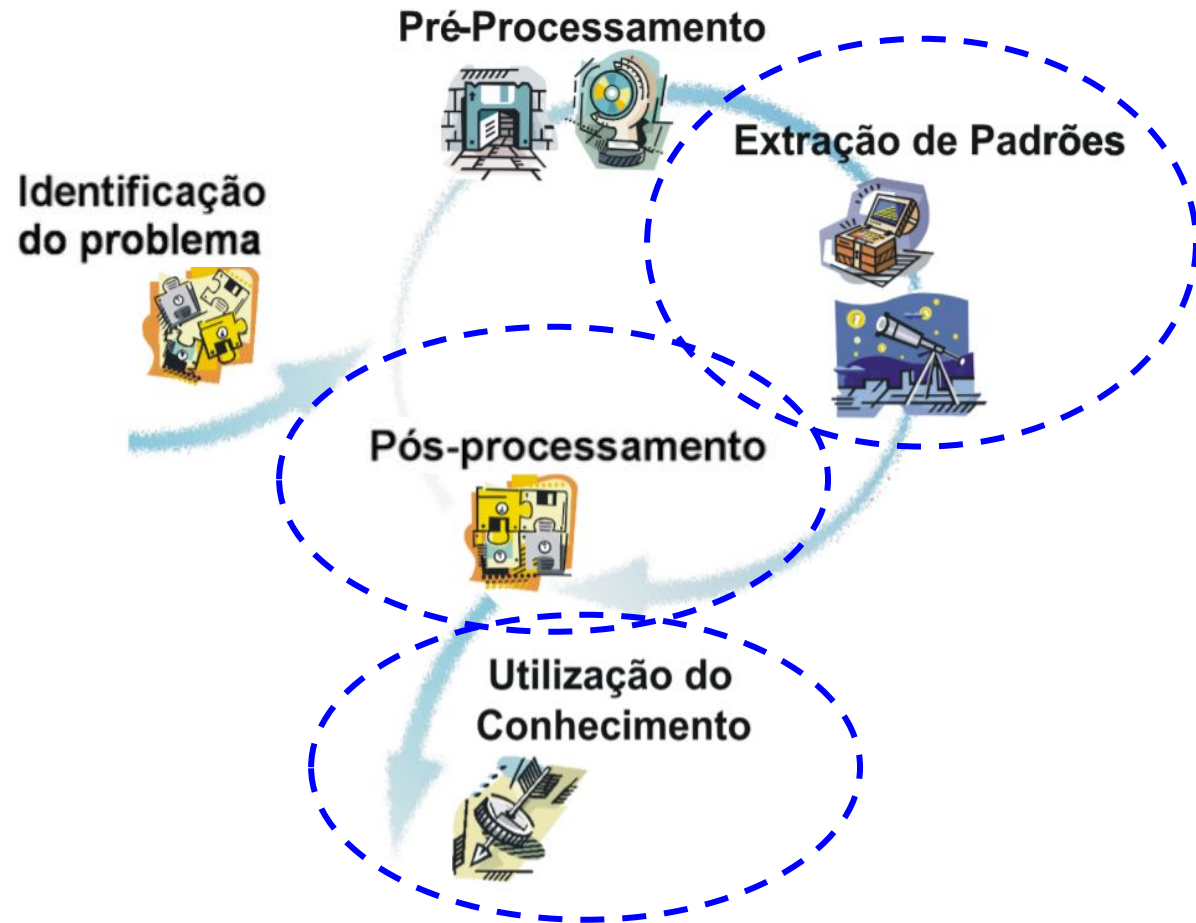
Regras de Associação

## Tarefas Preditivas

Classificação baseada em Instâncias



# Mineração de Dados



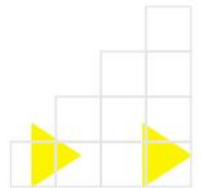
## Tarefas Descritivas

Agrupamento de Dados

Regras de Associação

## Tarefas Preditivas

Classificação baseada em Instâncias



# Bibliografia

Rezende, S. O. (2003). *Sistemas inteligentes: fundamentos e aplicações*. Editora Manole Ltda.

Tan, P.N.; Steinbach, M.; Karpatne, A.; Kumar, V. (2016). *Introduction to Data Mining (2nd Edition)*. Pearson.

