



Universidade Federal do Ceará
Campus de Sobral

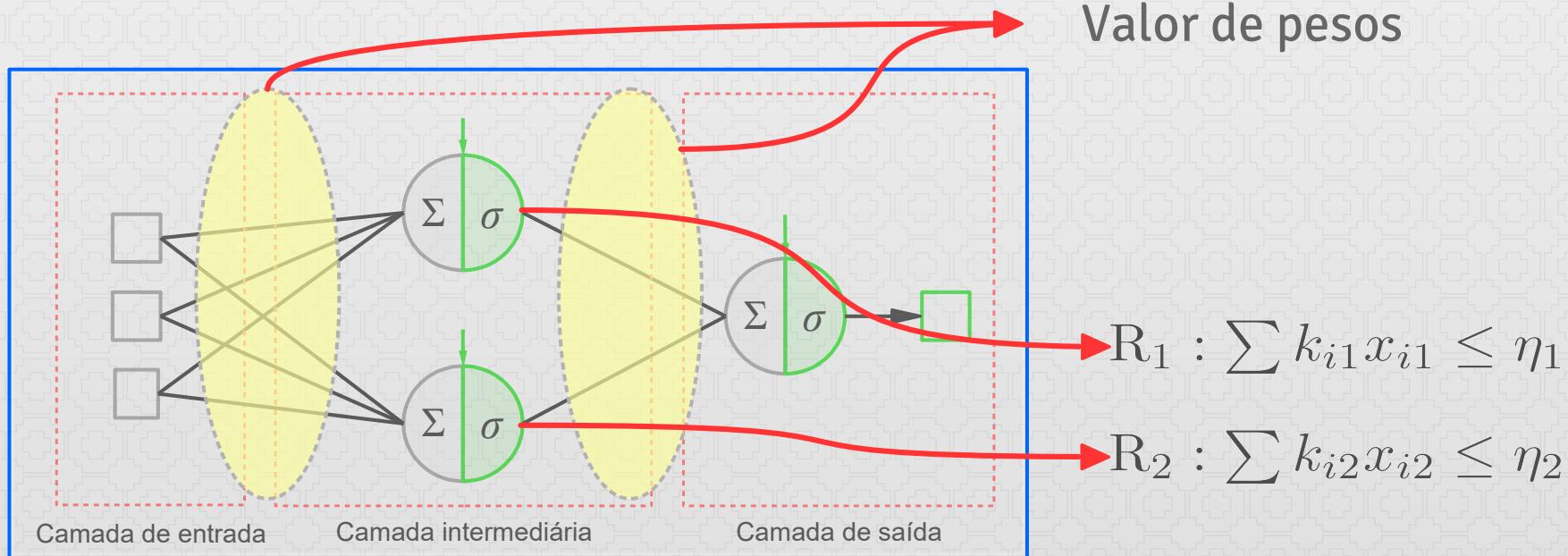
APLICAÇÃO DE UM ALGORITMO PARA EXTRAÇÃO DE REGRAS DE UMA REDE NEURAL ARTIFICIAL NO PROCESSO DE MINERAÇÃO DE DADOS

Ken Esparta Ccorahua

20 de dezembro de 2016

Contexto

- Extração de regras: Algoritmo RX.



Fonte: Autoria própria

Contexto

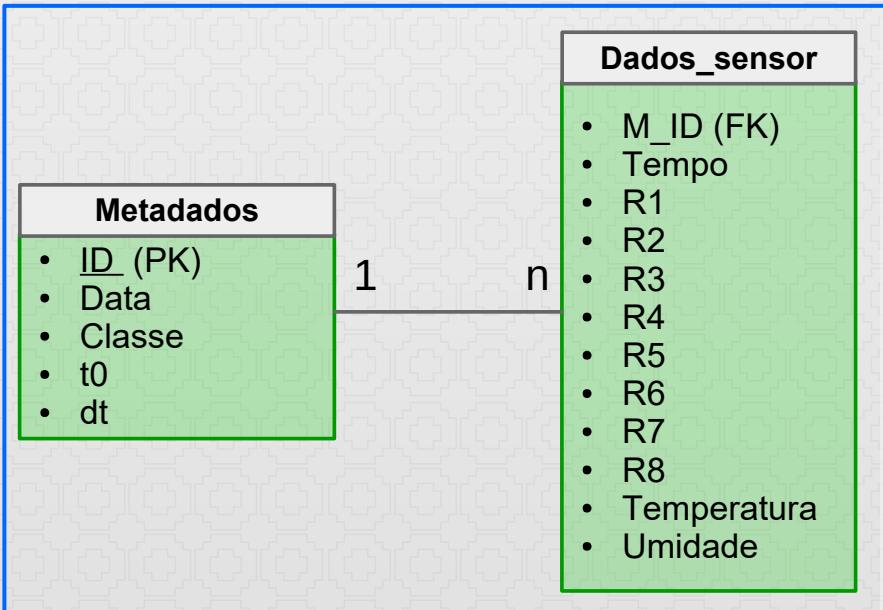
Gas sensors for home activity monitoring Data Set



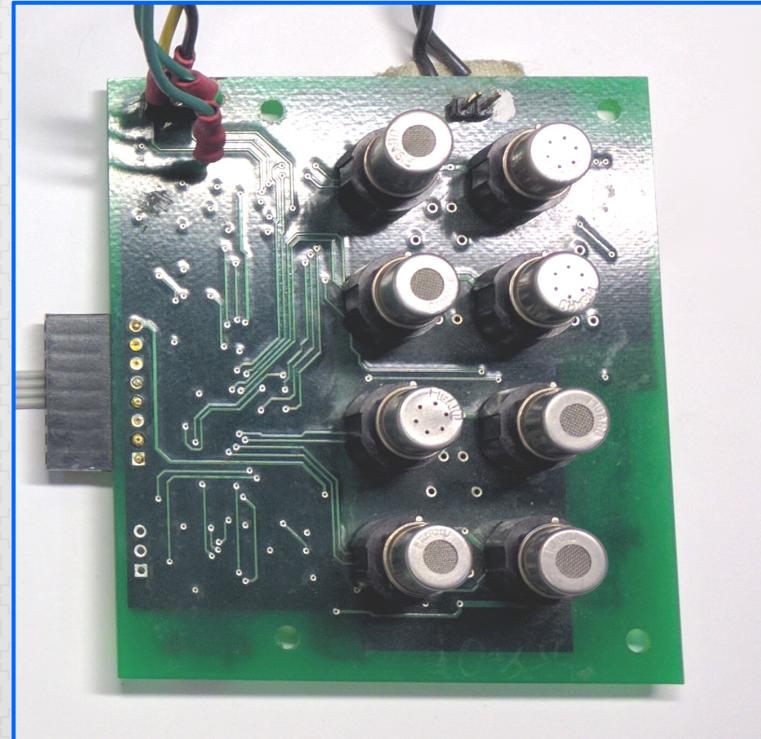
<https://archive.ics.uci.edu/ml/index.html>

Contexto

- Mineração de dados.



Fonte: (HUERTA et al., 2016)



Fonte: (HUERTA et al., 2016)

O problema

- Dificuldade de interpretação dos modelos da Rede Neural Artificial por parte de um especialista.
- Ausência de regras lógicas que cheguem a um resultado específico.

A solução

- Utilizar metodologias relacionadas com a extração de regras.
- As regras lógicas podem ser analisadas por um especialista, tirando a complexidade de uma rede neural artificial.

Objetivos

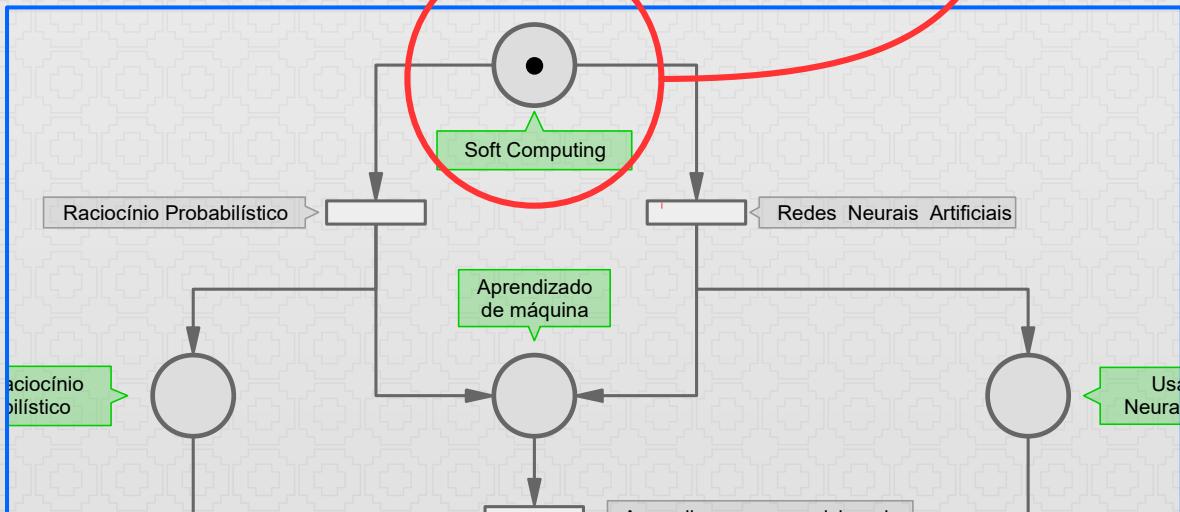
- Mostrar uma aplicação da extração de regras.
- Aplicar a metodologia CRISP-DM.
- Comparar o método de extração de regras com o método de árvore de decisão.

Justificativa

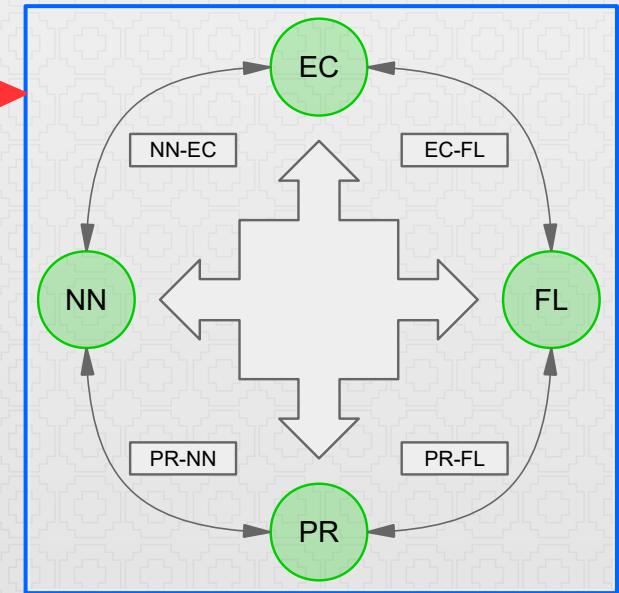
- Deseja-se extrair regras de uma rede neural artificial, utilizando o algoritmo RX, que possam ser interpretadas por um especialista.
- Mostrar uma situação prática de utilização da metodologia de mineração de dados CRISP-DM.

Fundamentação teórica

Soft Computing



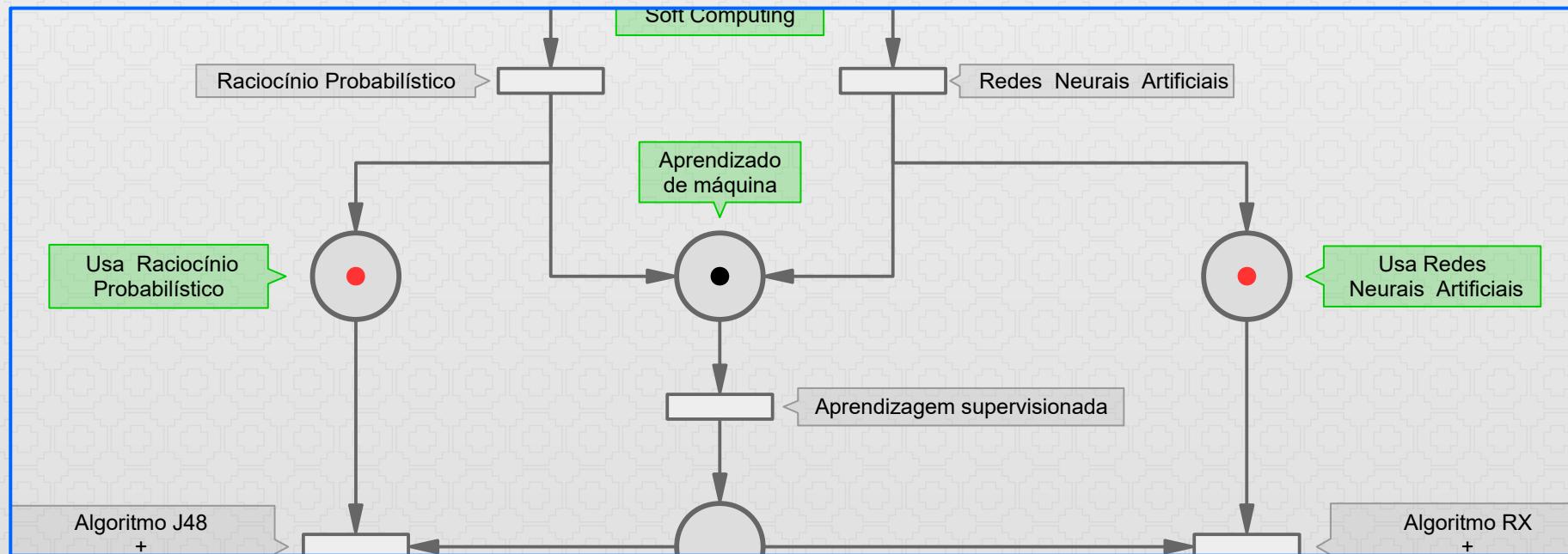
Fonte: Autoria própria



Fonte: (MANKAD, 2013).

Fundamentação teórica

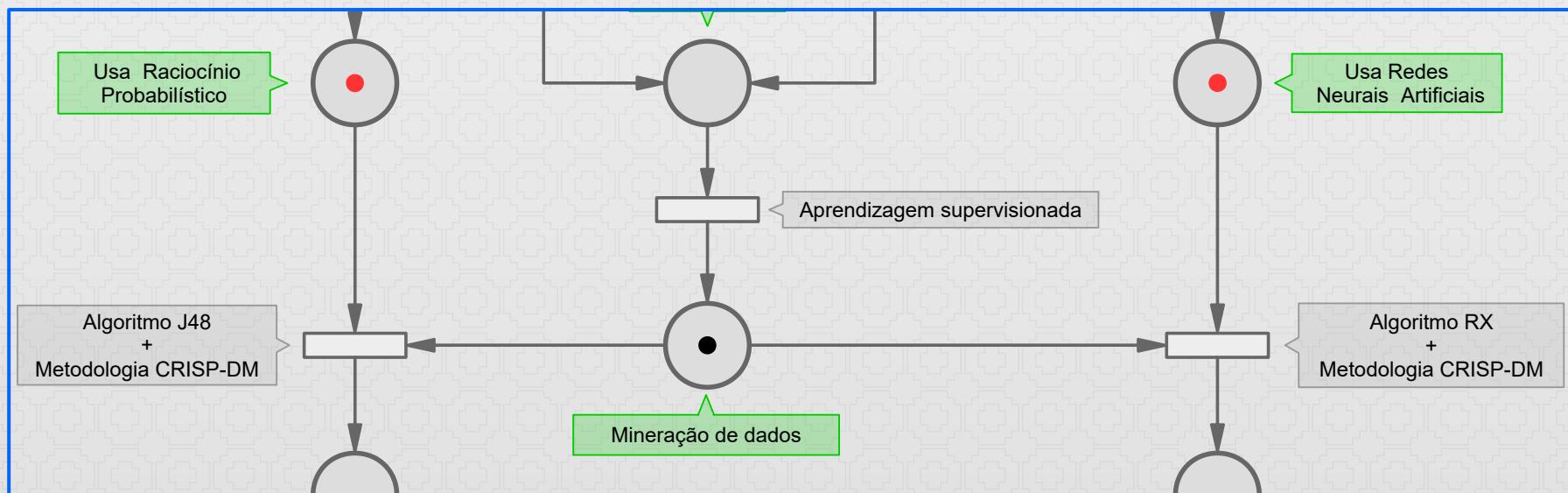
Aprendizado de máquina



Fonte: Autoria própria

Fundamentação teórica

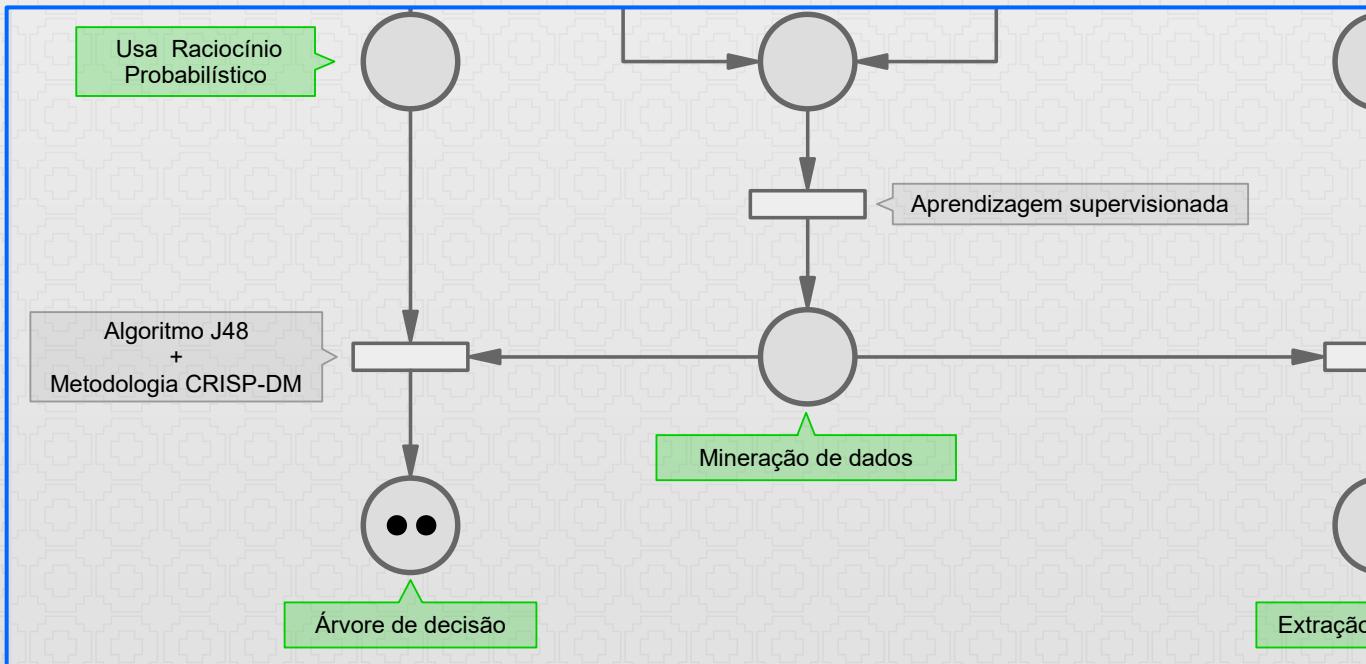
Mineração de dados



Fonte: Autoria própria

Fundamentação teórica

Árvore de decisão



Fonte: Autoria própria

Fundamentação teórica

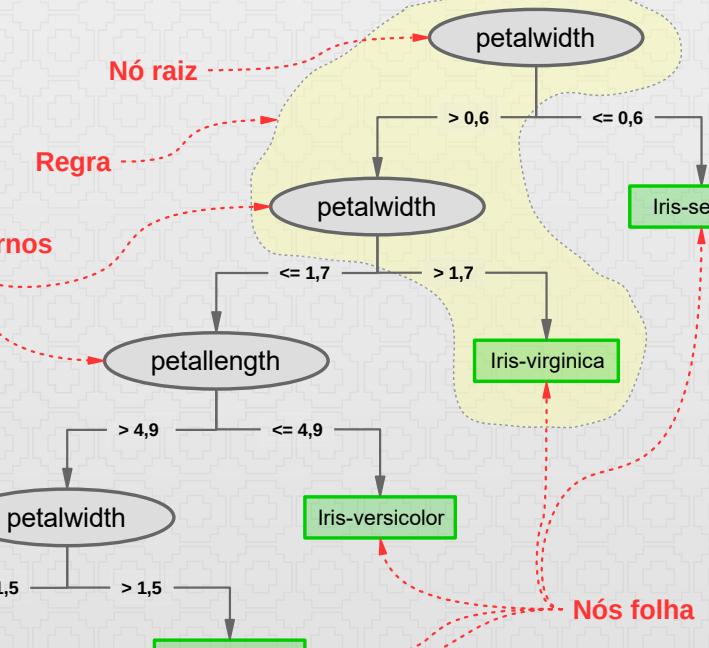
Árvore de decisão

```
petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6
    petalwidth <= 1.7
        petallength <= 4.9: Iris-versicolor (48.0/1.0)
        petallength > 4.9
            petalwidth <= 1.5: Iris-virginica (3.0)
            petalwidth > 1.5: Iris-versicolor (3.0/1.0)
    petalwidth > 1.7: Iris-virginica (46.0/1.0)
```



Fonte: Autoria própria

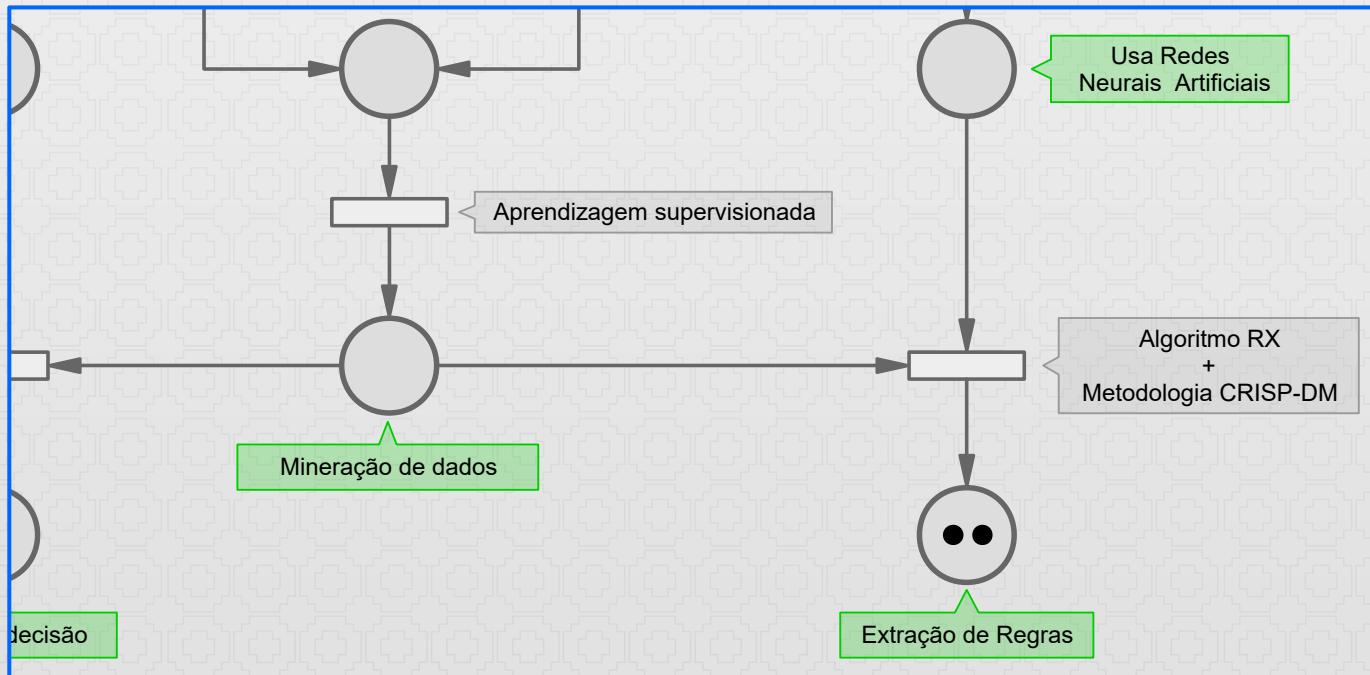
Diagrama da árvore de decisão



Fonte: Autoria própria

Fundamentação teórica

Extração de Regras



Fonte: Autoria própria

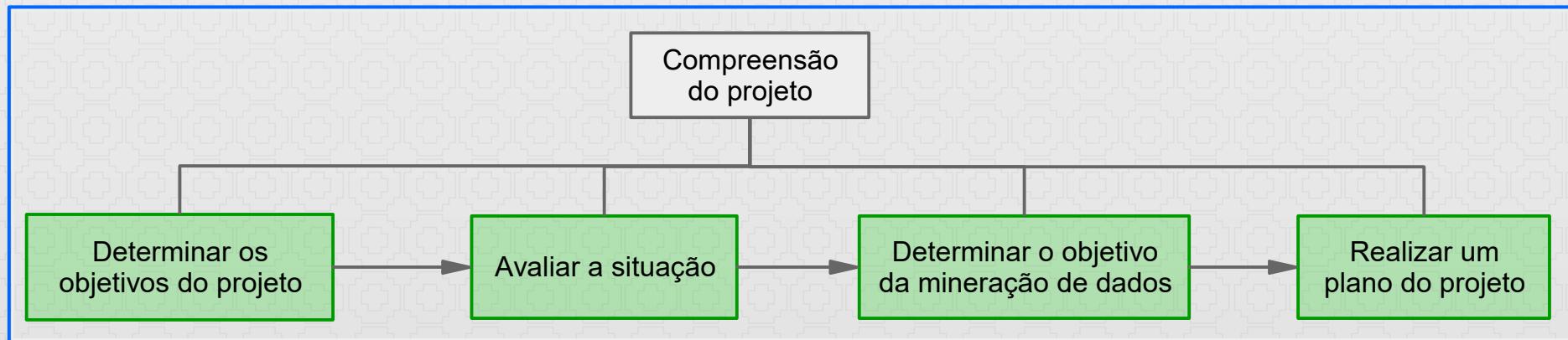
Fundamentação teórica

Extração de Regras: Algoritmo RX

- Treinar a rede neural.
- Utilizar o algoritmo de clustering.
- Gerar regras lógicas.
- Combinar as regras obtidas

A metodologia CRISP-DM

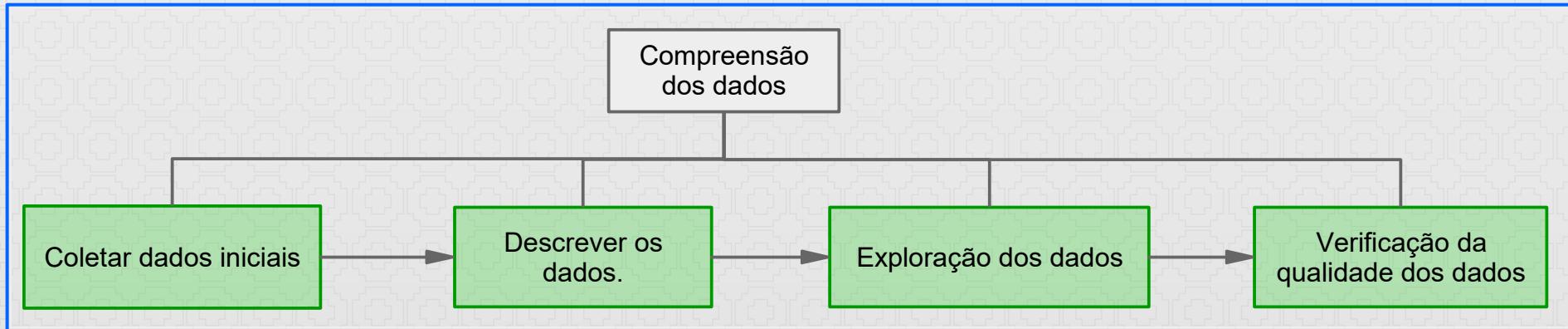
Compreensão do projeto



Fonte: (CHAPMAN et al., 2000)

A metodologia CRISP-DM

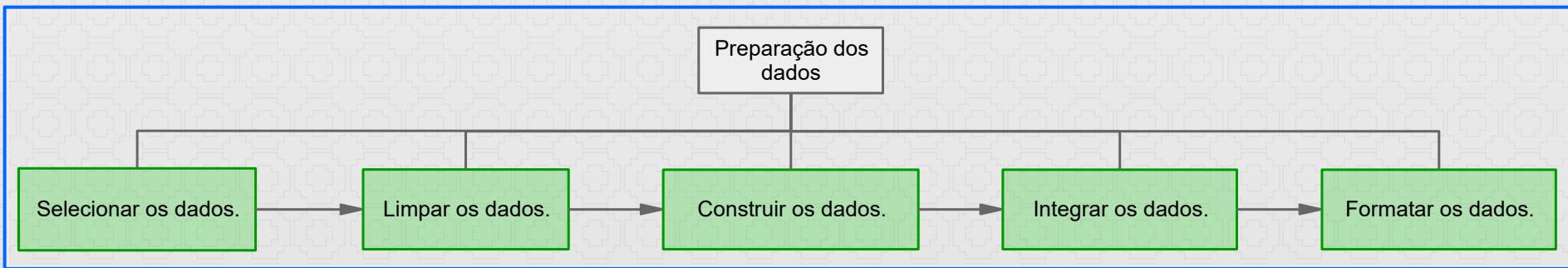
Compreensão dos dados



Fonte: (CHAPMAN et al., 2000)

A metodologia CRISP-DM

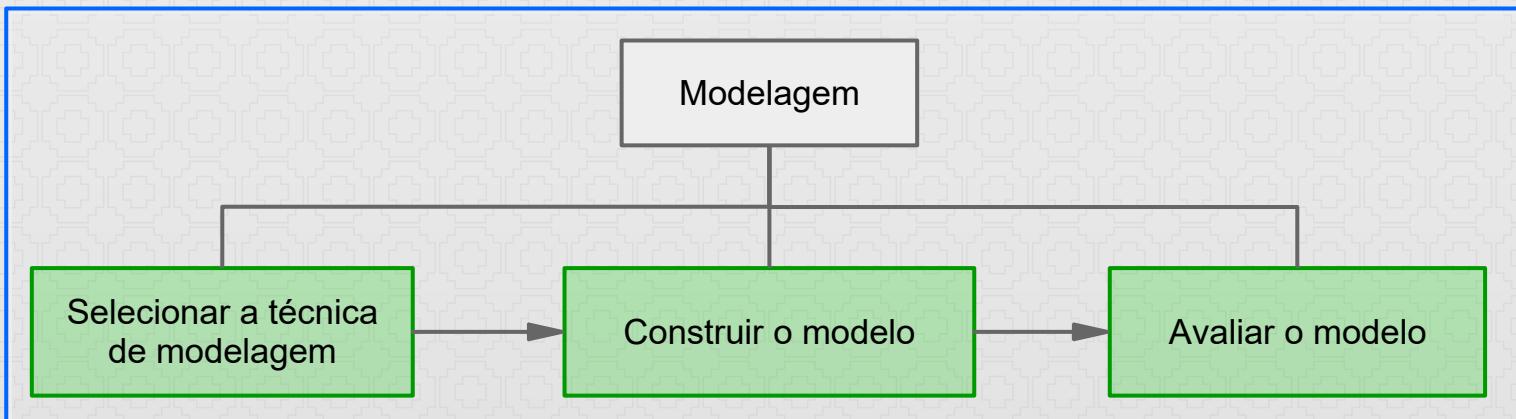
Preparação dos dados



Fonte: (CHAPMAN *et al.*, 2000)

A metodologia CRISP-DM

Modelagem



Fonte: (CHAPMAN *et al.*, 2000)

A metodologia CRISP-DM

Avaliação

- É importante avaliar o modelo final do projeto.
- Se reconhece o valor comercial.
- Ferramentas com as quais se mostram os resultados.

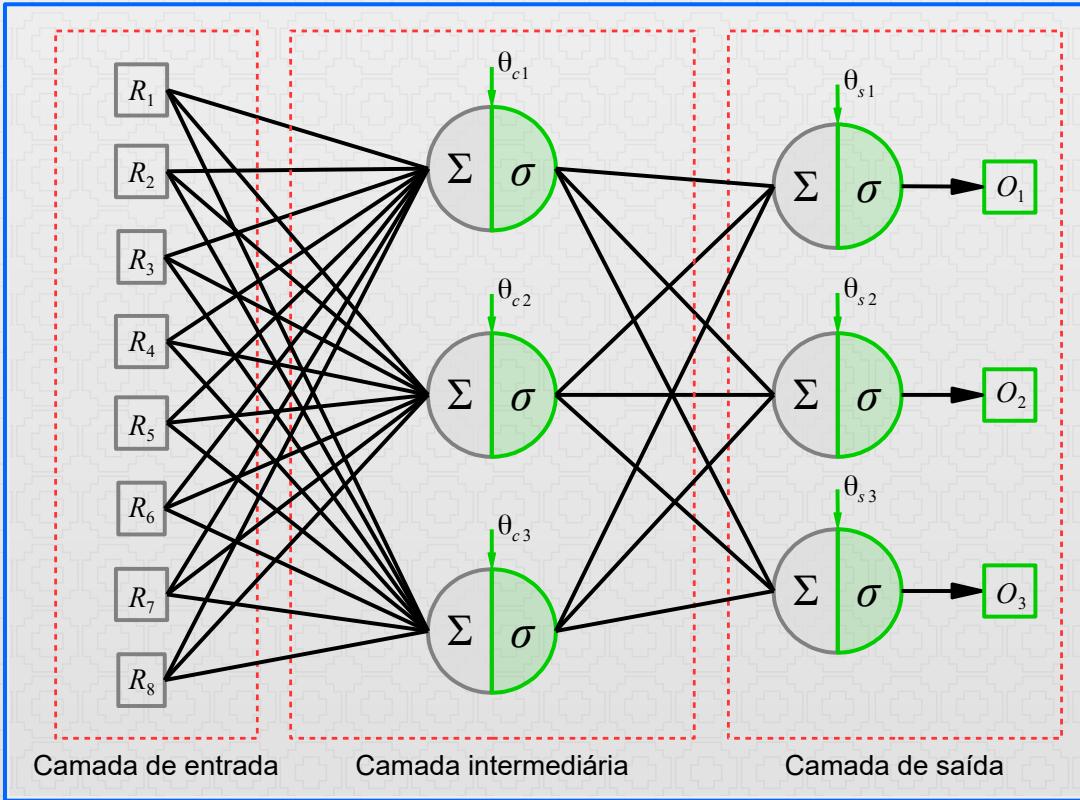
A metodologia CRISP-DM

Desenvolvimento

- Monitoramento do modelo proposto.
- Aumentar o conhecimento.

Resultados

Algoritmo RX



Fonte: Autoria própria

Resultados

Algoritmo RX

Equações de ativação

$$N_1 = 3.212 - 1.101R_1 + 0.154R_2 - 0.933R_3 + 1.886R_4 + \dots \\ \dots + 1.395R_5 + 1.610R_6 - 0.315R_7 - 0.135R_8$$

$$N_2 = 0.000 - 0.132R_1 - 1.859R_2 - 1.156R_3 + 1.183R_4 + \dots \\ \dots + 0.796R_5 - 0.489R_6 - 1.723R_7 + 0.521R_8$$

$$N_3 = -3.212 - 0.793R_1 + 1.120x2 + 0.802R_3 - 1.665R_4 + \dots \\ \dots + 0.883R_5 - 1.646R_6 - 0.771R_7 + 0.969R_8$$

Resultados

Algoritmo RX

Regras oblíquas (Banana)

SE $3.07 \leq N_1 \leq 3.74$ E
 $-2.22 \leq N_2 \leq -0.91$ E
 $-4.03 \leq N_3 \leq -3.65$ ENTÃO banana

Resultados

Algoritmo RX

Regras oblíquas (vinho)

SE $2.98 \leq N_1 \leq 3.83$ E
 $-2.27 \leq N_2 \leq -0.63$ E
 $-4.18 \leq N_3 \leq -3.57$ ENTÃO vinho

Resultados

Algoritmo RX

Regras oblíquas (background)

SE $3.33 \leq N_1 \leq 3.60$ E

$-1.98 \leq N_2 \leq -1.49$ E

$-3.99 \leq N_3 \leq -3.80$ ENTÃO background

Resultados

Algoritmo RX

Classificação correta: 89,78%

Resultados

Algoritmo J48

Classificação correta: 93,40%

Número de folhas da árvore é 217.

Tamanho da árvore é de 432.

```
1 | r6 <= 0.132509
2 |   | r7 <= 0.021227: vinho (2066.0/30.0)
3 |   | r7 > 0.021227
4 |   |   | r8 <= 0.068634
5 |   |   | r8 <= 0.06017
6 |   |   |   | r6 <= 0.098656
7 |   |   |   | r5 <= 0.024898
8 |   |   |   |   | r3 <= 0.366412: vinho (31.0)
9 |   |   |   |   | r3 > 0.366412
10 |   |   |   |   |   | r1 <= 0.746925
11 |   |   |   |   |   | r6 <= 0.078162
12 |   |   |   |   |   |   | r4 <= 0.496367: banana (117.0/3.0)
13 |   |   |   |   |   |   | r4 > 0.496367: vinho (50.0/8.0)
14 |   |   |   |   |   |   | r6 > 0.078162: banana (840.0/12.0)
```

Fonte: Autoria própria

Conclusão

- A extração de regras adaptou-se bem no processo.
- A metodologia CRISP-DM tem ferramentas que constroem uma base sólida.
- As regras obtidas pelo algoritmo RX são menores do que as obtidas pelo algoritmo J48.

Recomendações e trabalhos futuros

- Utilizar diferentes métodos.
- Aplicar em um experimento em tempo real.
- Desenvolver uma aplicação que mostre os resultados.

Obrigado!

Obrigado!!!
Muito