

Processo de Descoberta de Conhecimento

Fabrício J. Barth

Processo para a criação de modelos

- * Independentemente se o objetivo final de um **projeto de aprendizagem de máquina** é o desenvolvimento de um modelo **preditivo** (aprendizagem de máquina supervisionada) ou o desenvolvimento de um modelo **descritivo** (aprendizagem de máquina não supervisionada), em ambos os casos, é desejável **seguir** um **processo** para que este trabalho possa ser executado de uma forma mais **eficiente**.
- * O objetivo deste material é apresentar um processo chamado **“Processo de Descoberta de Conhecimento”** que pode ser aplicado no desenvolvimento de projetos que envolvam aprendizagem de máquina.

Processo de Descoberta de Conhecimento (KDD - Knowledge Discovery in Databases)

Qual é a pergunta?

É possível classificar espécies do gênero iris levando em consideração apenas o tamanho das plantas?



Iris virginica



Iris setosa



Iris versicolor



Processo de Descoberta de Conhecimento (KDD - Knowledge Discovery in Databases)

Qual é a pergunta?

Aquisição e pré-processamento dos dados

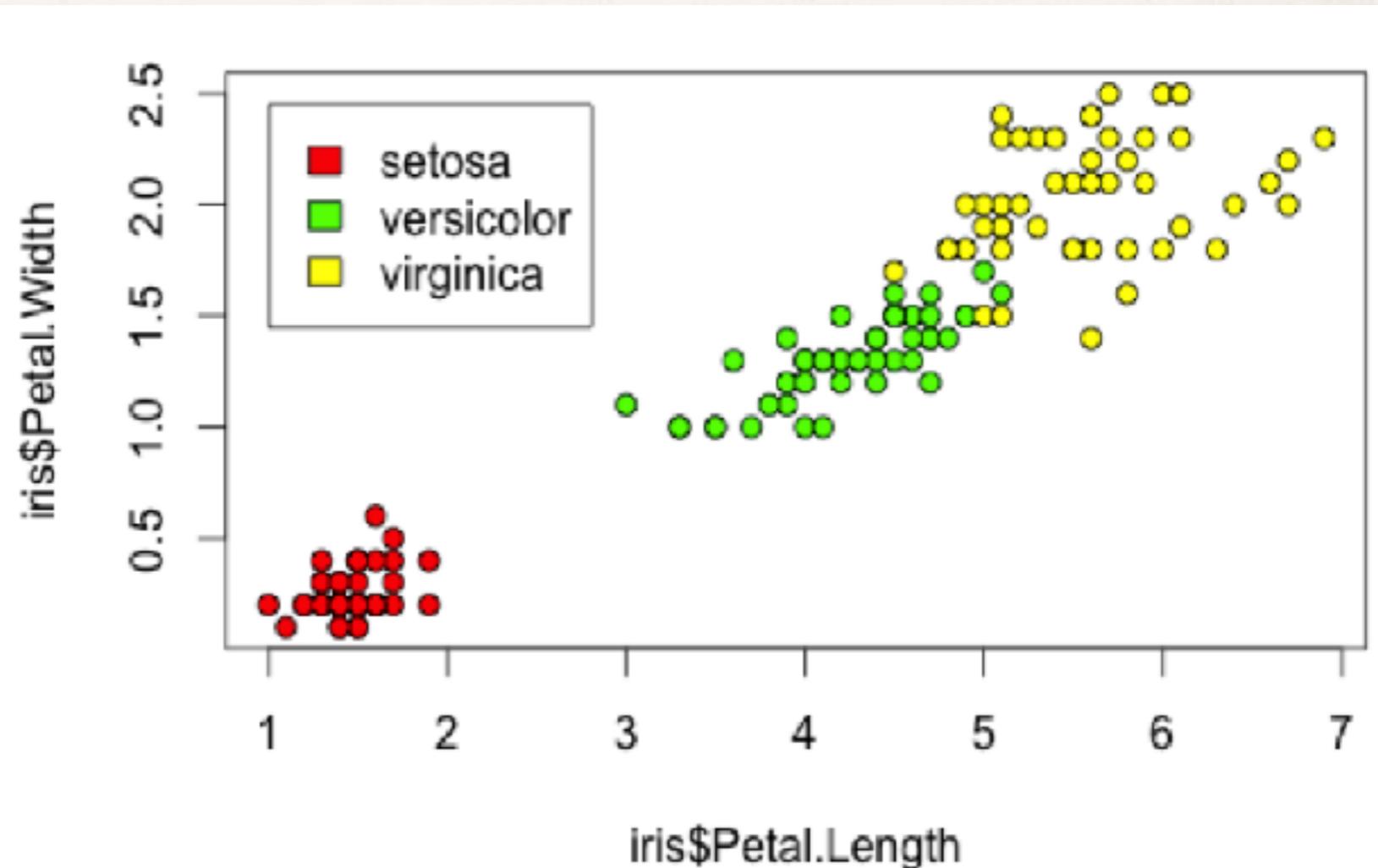
```
> data(iris)
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1           5.1         3.5         1.4         0.2  setosa
2           4.9         3.0         1.4         0.2  setosa
3           4.7         3.2         1.3         0.2  setosa
4           4.6         3.1         1.5         0.2  setosa
5           5.0         3.6         1.4         0.2  setosa
6           5.4         3.9         1.7         0.4  setosa
>
>
> sapply(iris,class)
Sepal.Length Sepal.Width Petal.Length Petal.Width Species
  "numeric"   "numeric"   "numeric"   "numeric"   "factor"
```

Processo de Descoberta de Conhecimento (KDD - Knowledge Discovery in Databases)

Qual é a pergunta?

Aquisição e pré-processamento dos dados

Análise exploratória



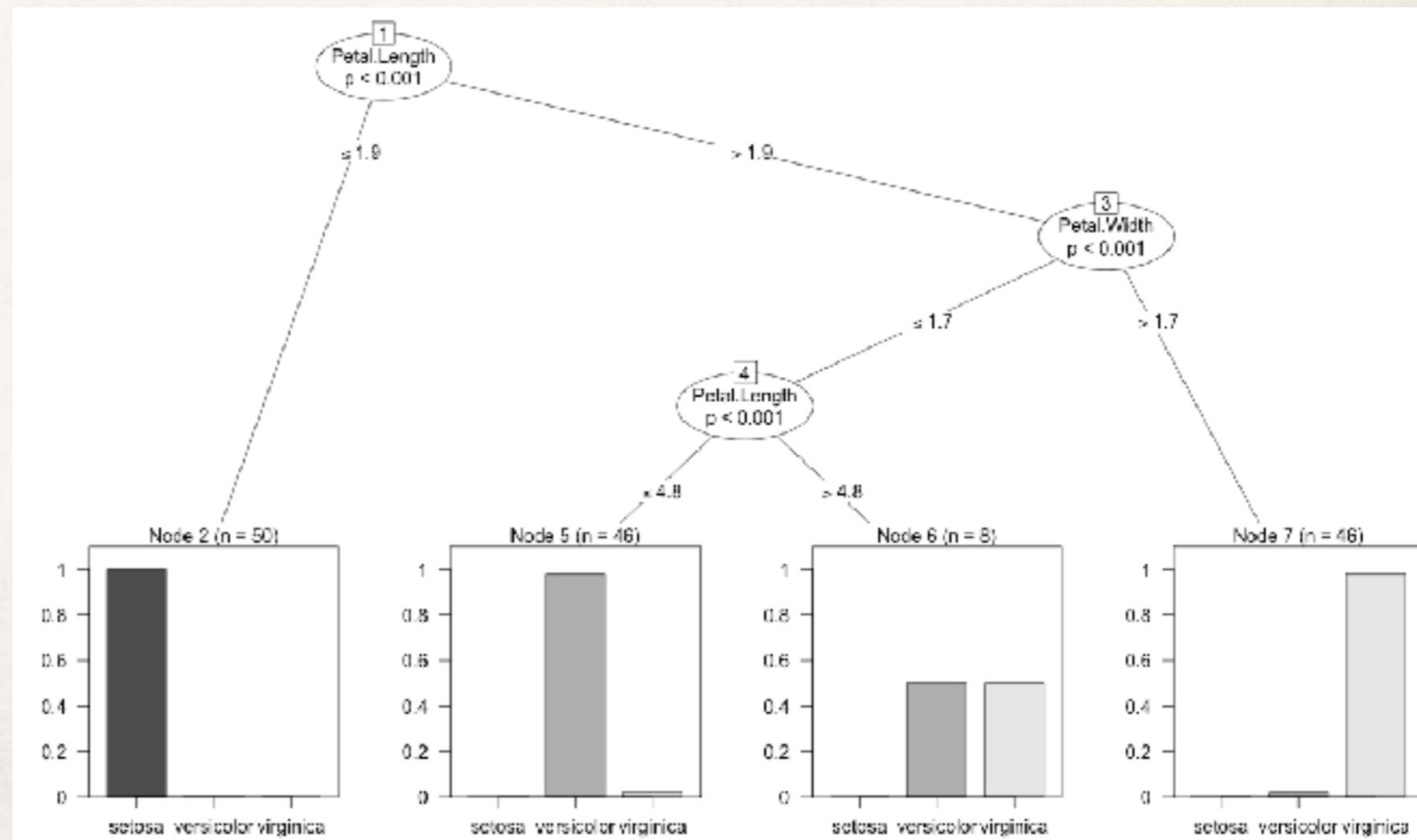
Processo de Descoberta de Conhecimento (KDD - Knowledge Discovery in Databases)

Qual é a pergunta?

Aquisição e pré-processamento dos dados

Análise exploratória

Modelagem



Processo de Descoberta de Conhecimento (KDD - Knowledge Discovery in Databases)

Qual é a pergunta?

Aquisição e pré-processamento dos dados

Análise exploratória

Modelagem

Avaliação do modelo

```
> table(predict(model,iris), iris$Species)
```

	setosa	versicolor	virginica
setosa	50	0	0
versicolor	0	49	5
virginica	0	1	45

Acurácia do modelo?

Quantidade falsos positivos?

Falsos negativos?

Processo de Descoberta de Conhecimento (KDD - Knowledge Discovery in Databases)

Qual é a pergunta?

Aquisição e pré-processamento dos dados

Análise exploratória

Modelagem

Avaliação do modelo

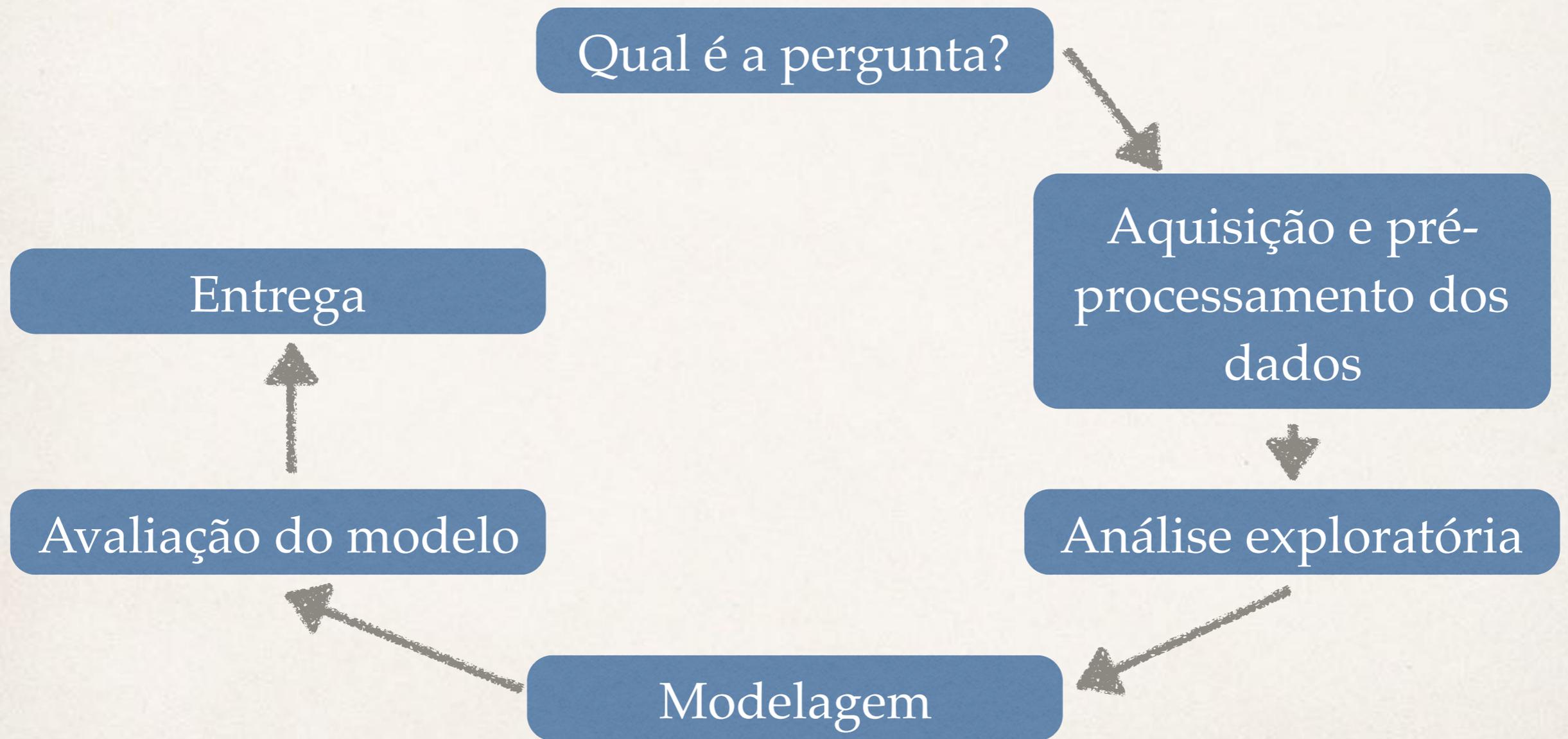
Entrega

Relatórios Estáticos

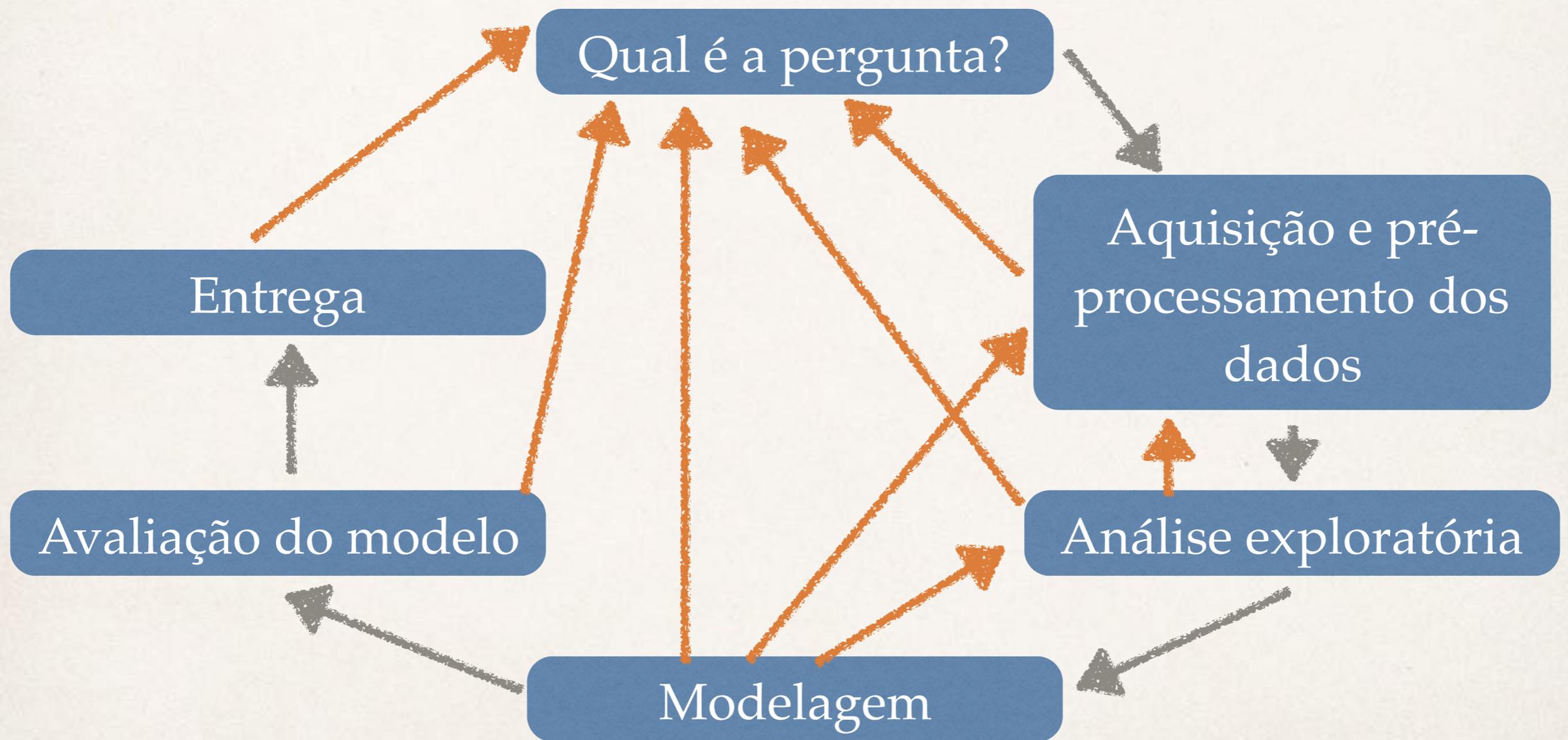
Relatórios Dinâmicos

Aplicativos

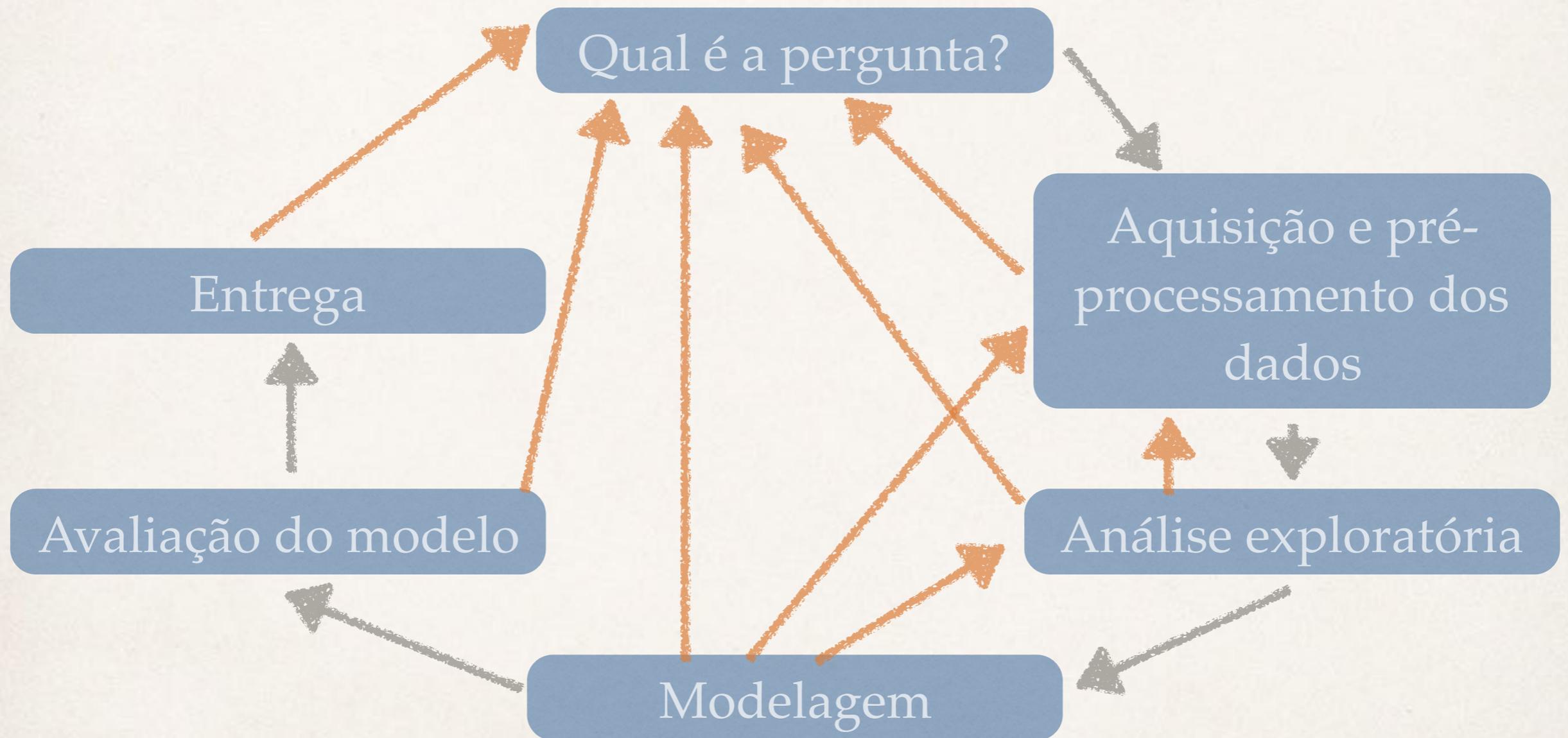
Processo de Descoberta de Conhecimento (KDD - Knowledge Discovery in Databases)



Processo de Descoberta de Conhecimento (KDD - Knowledge Discovery in Databases)



Processo de Descoberta de Conhecimento (KDD - Knowledge Discovery in Databases)



Este processo pode ser suportado por diversas ferramentas, entre elas: R, SPSS, RapidMiner, Tableau, Weka, Matlab, Octave, Python, Julia,...

Referência

- * Fayyad, Piatetsky-Shapiro, Smyth, "From Data Mining to Knowledge Discovery: An Overview", in Fayyad, Piatetsky-Shapiro, Smyth, Uthurusamy, *Advances in Knowledge Discovery and Data Mining*, AAAI Press / The MIT Press, Menlo Park, CA, 1996, pp.1-34