

# Formação Cientista de Dados

Classificação

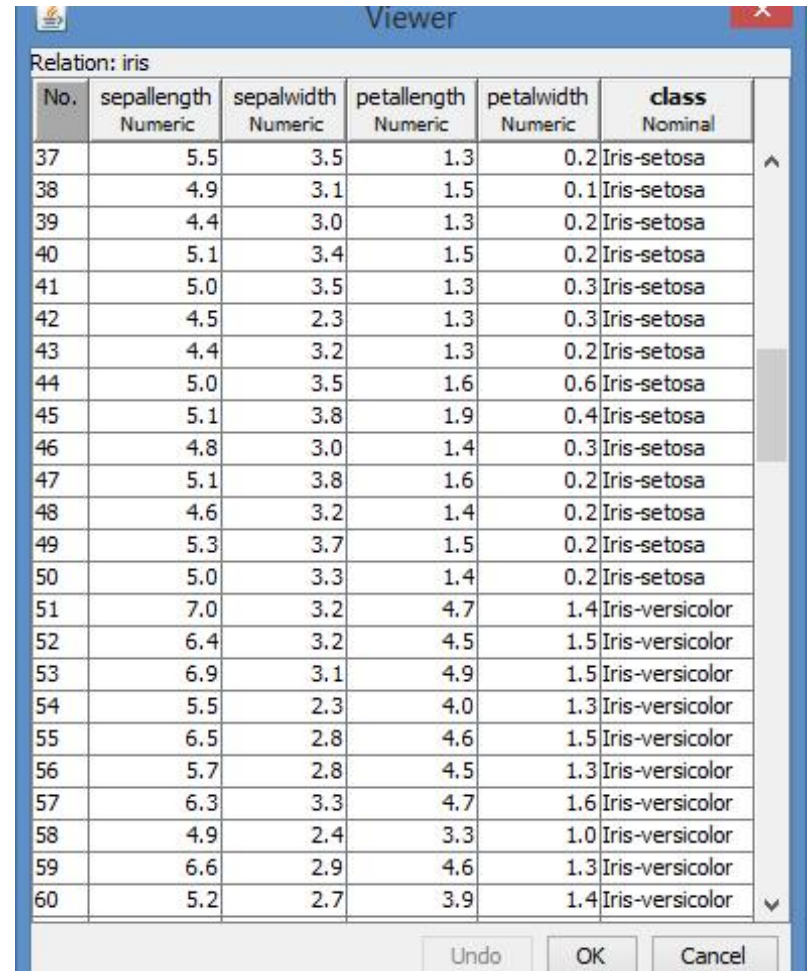
# Classificação

- Queremos descobrir ou descrever a classe de um fato.
- Normalmente a classe em uma relação esta representada em um atributo especial, posicionado como último atributo da relação



## Relação Íris

- 150 instâncias e
- 5 atributos.
- 4 atributos numéricos
- 1 atributo nominal, Species, que é a classe: setosa, versicolor, virginica



Viewer

Relation: iris

No.	sepallength Numeric	sepalwidth Numeric	petallength Numeric	petalwidth Numeric	class Nominal
37	5.5	3.5	1.3	0.2	Iris-setosa
38	4.9	3.1	1.5	0.1	Iris-setosa
39	4.4	3.0	1.3	0.2	Iris-setosa
40	5.1	3.4	1.5	0.2	Iris-setosa
41	5.0	3.5	1.3	0.3	Iris-setosa
42	4.5	2.3	1.3	0.3	Iris-setosa
43	4.4	3.2	1.3	0.2	Iris-setosa
44	5.0	3.5	1.6	0.6	Iris-setosa
45	5.1	3.8	1.9	0.4	Iris-setosa
46	4.8	3.0	1.4	0.3	Iris-setosa
47	5.1	3.8	1.6	0.2	Iris-setosa
48	4.6	3.2	1.4	0.2	Iris-setosa
49	5.3	3.7	1.5	0.2	Iris-setosa
50	5.0	3.3	1.4	0.2	Iris-setosa
51	7.0	3.2	4.7	1.4	Iris-versicolor
52	6.4	3.2	4.5	1.5	Iris-versicolor
53	6.9	3.1	4.9	1.5	Iris-versicolor
54	5.5	2.3	4.0	1.3	Iris-versicolor
55	6.5	2.8	4.6	1.5	Iris-versicolor
56	5.7	2.8	4.5	1.3	Iris-versicolor
57	6.3	3.3	4.7	1.6	Iris-versicolor
58	4.9	2.4	3.3	1.0	Iris-versicolor
59	6.6	2.9	4.6	1.3	Iris-versicolor
60	5.2	2.7	3.9	1.4	Iris-versicolor

Undo OK Cancel

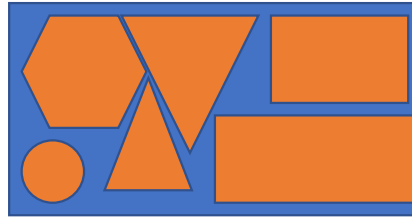
# Outros exemplos

Arquivo ARFF	Descrição	Atributos	Classe
breast-cancer	Dados de pacientes com câncer de mama	Idade, menopausa, tamanho do tumor, etc.	Se os eventos são recorrentes ou não
contact-lenses	Tipos de lentes de contatos que melhor se ajustam em pacientes	Idade, astigmatismo, produção de lagrima	Se o paciente deve receber lentes de contato macias, duras ou se não deve receber lentes
credit-g	Dados de solicitação de crédito da Alemanha	Tempo de uso de cheque, histórico de credito, propósito etc.	Se o cliente é um bom ou mal pagador de empréstimos
iris	Dados de medidas de sépala e pétala de flores Iris	Largura e comprimento da pétala e da sépala	A espécie da classe: setosa, versicolor, virginica
Soybean	Dados de plantações de soja	Diversas características da plantação, como temperatura, germinação, folhas etc.	Diversos tipos de doenças de soja
Wather	Condições do tempo	Aparência, temperatura, umidade, vento	Se é ou não possível jogar



# Aprendizado de Máquina

- Algoritmo I: calcular o aproveitamento do papel para uma gráfica



- Algoritmo II: prever se pessoa será boa ou má pagadora de um empréstimo



# Aprendizado de Máquina

Gráfica	Crédito
Baseado em Entrada: Dados atuais	Baseado em Dados Históricos
Algoritmo puro	Algoritmo + Modelo
100% performance	Não se espera 100%
Performance constante	Performance varia
Atende qualquer negócio	Adequa-se ao negócio
Não precisa aprender	Precisa aprender e reaprender

---

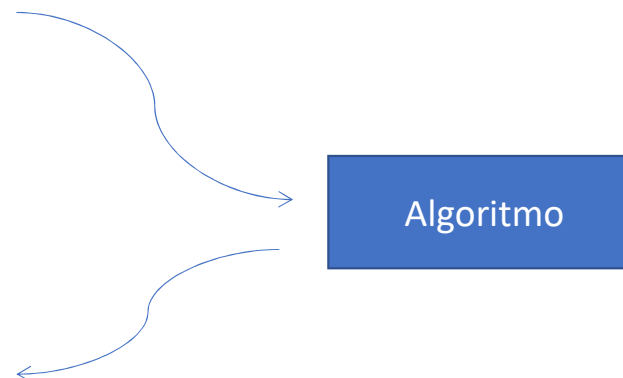
# Modelo

## Dados Históricos de Concessão de Crédito

Idade	Pagou
18	Não
46	Sim
34	Sim
21	Não
37	Não
...	

## Modelo Construído

Idade	Bom Pagador
18~22	Não
23~35	Sim
36~45	Não
45~65	Sim

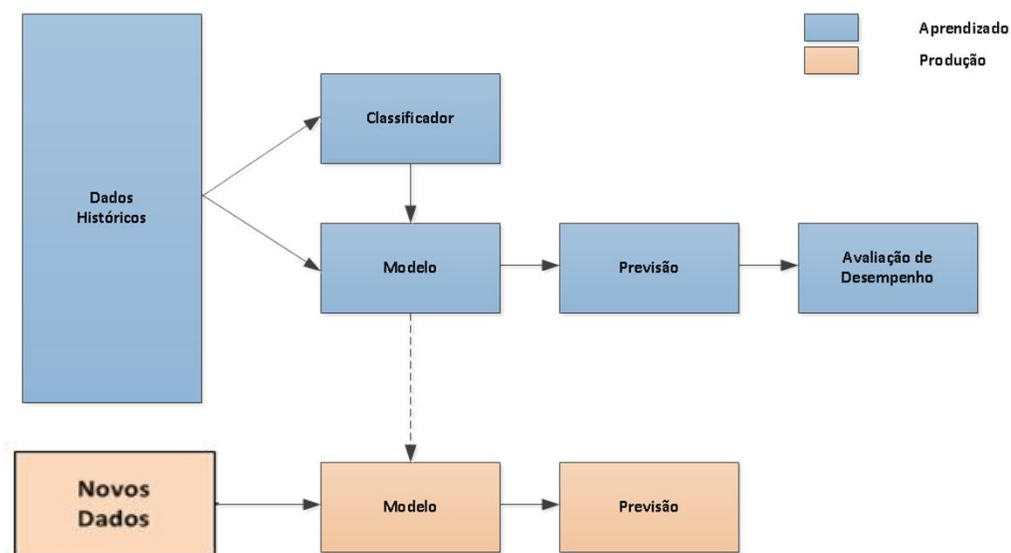


Novo cliente com 37 anos. Bom ou mal pagador?



# Avaliando o que foi aprendido

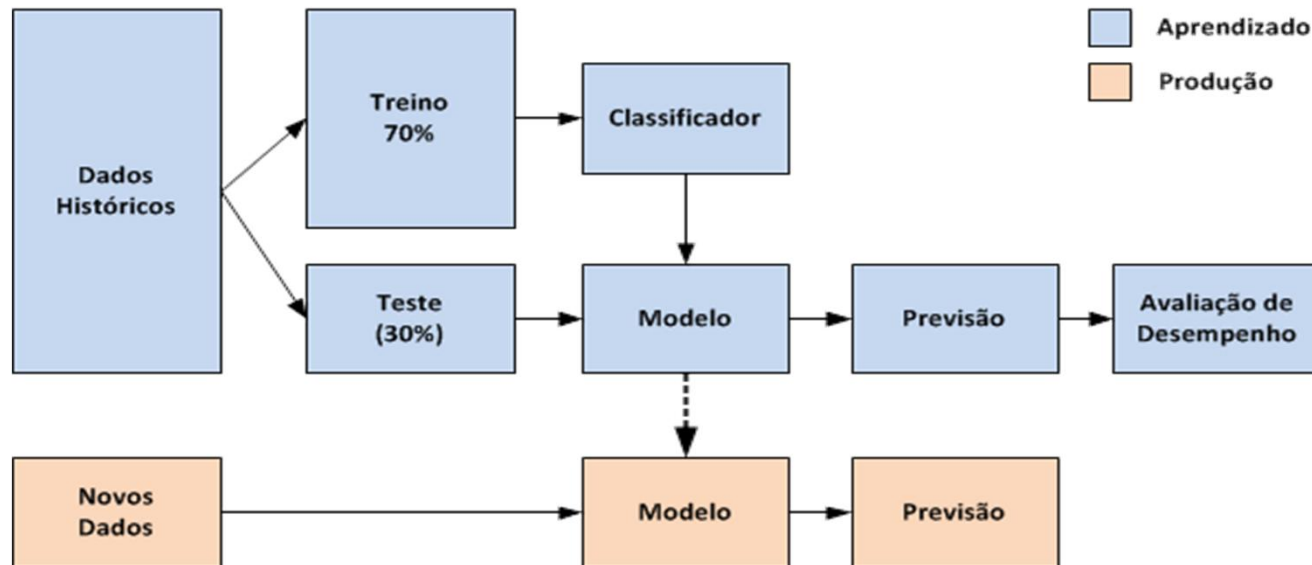
## 1 - Usando mesmo conjunto de dados





# Avaliando o que foi aprendido

## 2 – Hold out



# Avaliando o que foi aprendido

## 3 – Validação Cruzada

