

FORMAÇÃO CIENTISTA DE DADOS

CLASSIFICAÇÃO

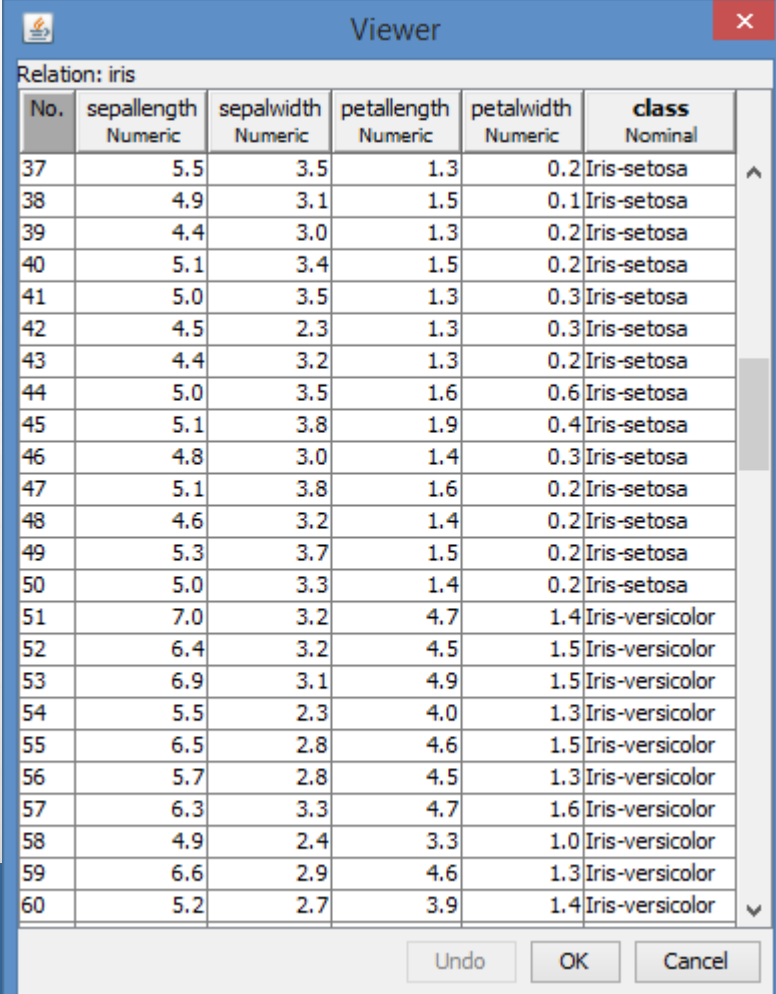


CLASSIFICAÇÃO

- **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 IRIS

- **150 INSTÂNCIAS E**
- **5 ATRIBUTOS.**
- **4 ATRIBUTOS NUMÉRICOS**
- **1 ATRIBUTO NOMINAL, SPECIES, QUE É A CLASSE: SETOSA, VERSICOLOR, VIRGINICA**



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.arff	Dados de pacientes com câncer de mama	Idade, menopausa, tamanho do tumor, etc.	Se os eventos são recorrentes ou não
contact-lenses.arff	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.arff	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.arff	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.arff	Dados de plantações de soja	Diversas características da plantação, como temperatura, germinação, folhas etc.	Diversos tipos de doenças de soja
Wather.arff	Condições do tempo	Aparência, temperatura, umidade, vento	Se é ou não possível jogar

Relation: german_credit

No.	checking_status Nominal	duration Numeric	credit_history Nominal	purpose Nominal	credit_amount Numeric	savings_status Nominal	employment Nominal
30	(0)	60.0	delayed previously	business	6836.0	(100)	=7

Navigation: < [] >

Actions: Undo OK Cancel

APRENDIZADO DE MAQUINA

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

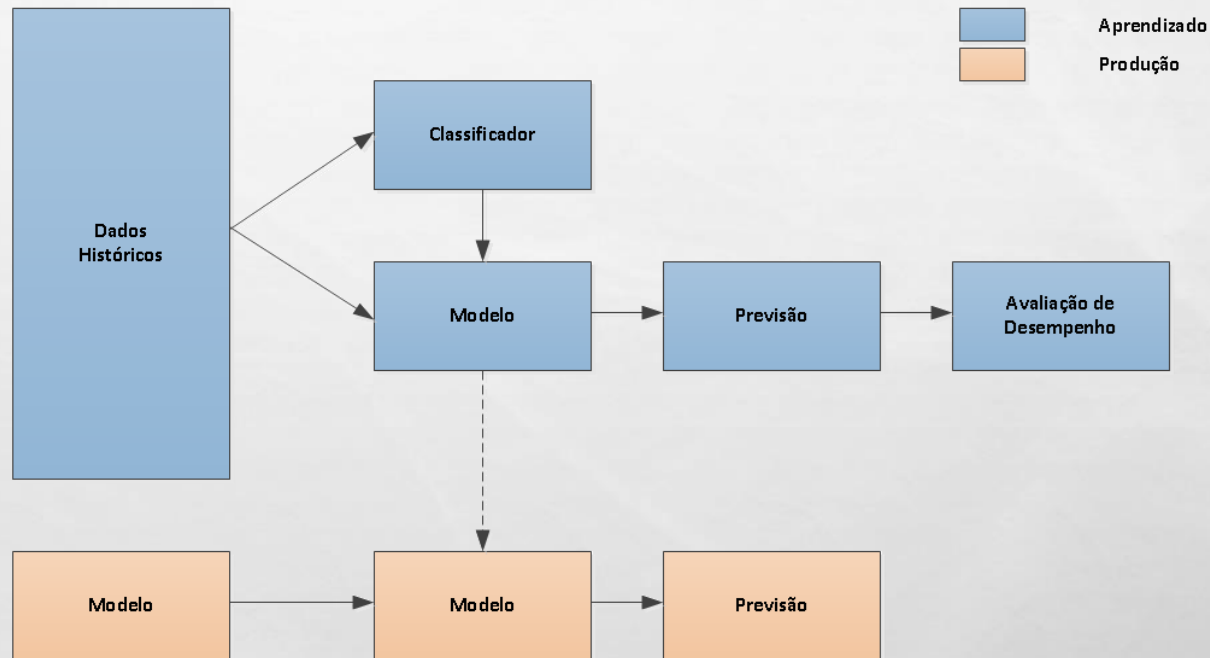


Algorithm

Novo cliente com 37 anos. Bom ou mal pagador?

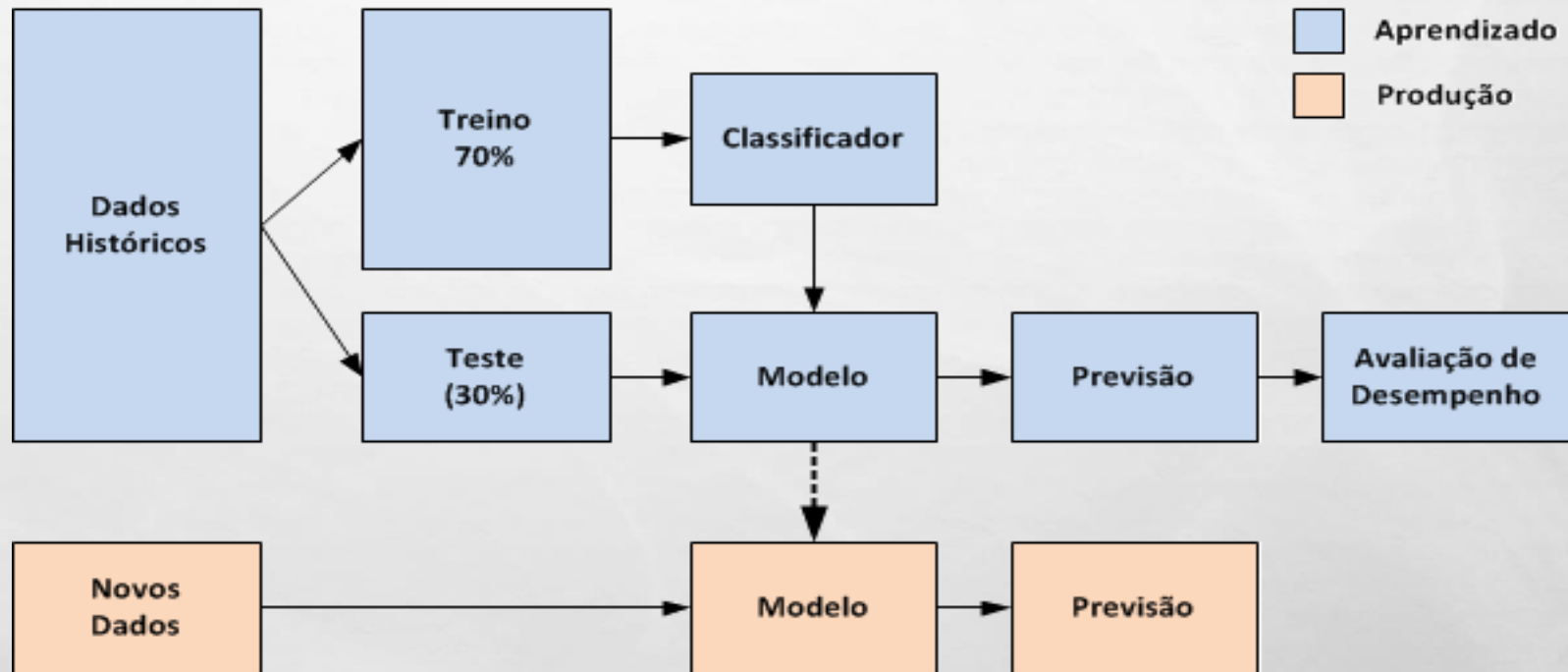
AVALIANDO O QUE FOI APRENDIDO

- 1 - USANDO MESMO CONJUNTO DE DADOS



AVALIANDO O QUE FOI APRENDIDO

- **2 – HOLD OUT**



AVALIANDO O QUE FOI APRENDIDO

- 3 – VALIDAÇÃO CRUZADA

