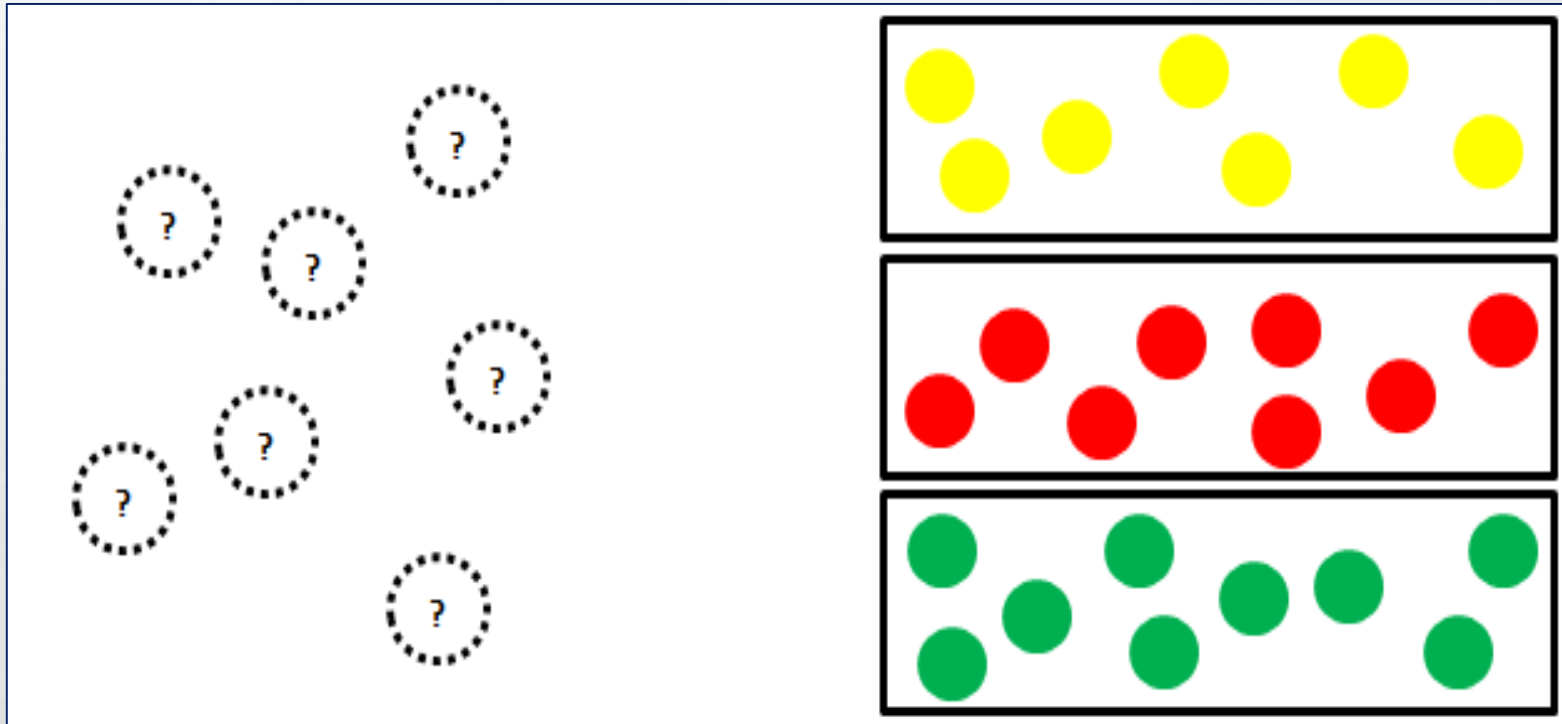


# FORMAÇÃO CIENTISTA DE DADOS

**MACHINE LEARNING: CONCEITOS**

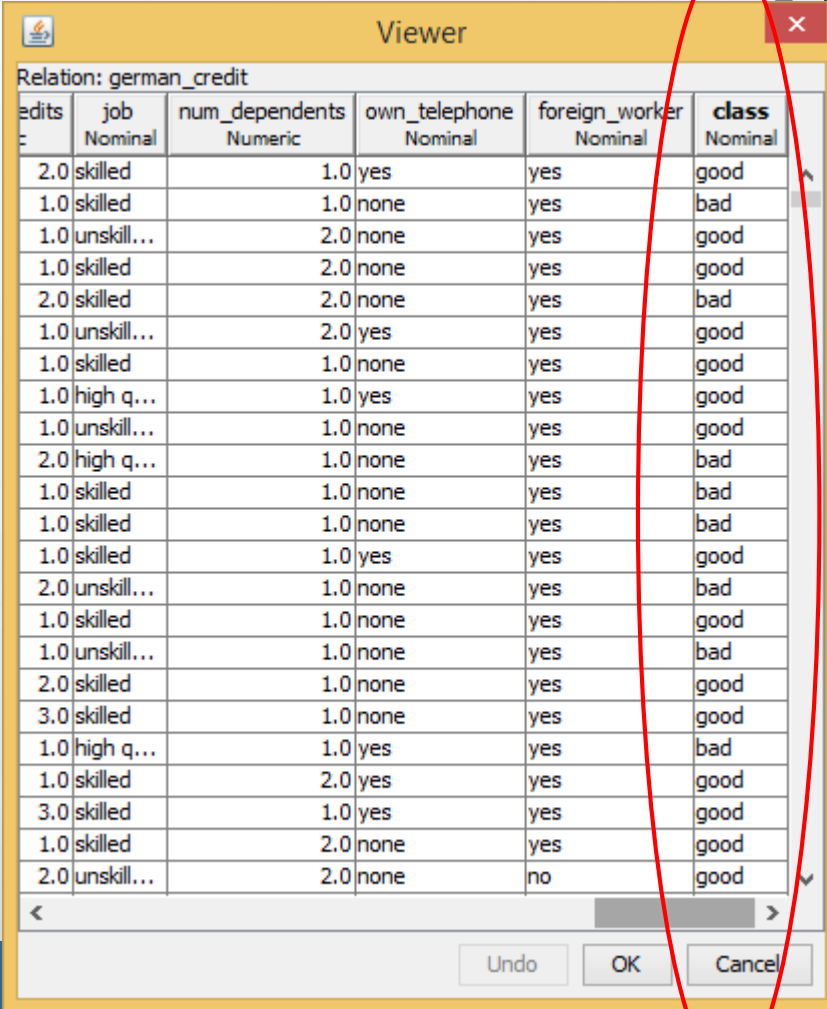


# CLASSIFICAÇÃO



# CLASSIFICAÇÃO

- **DESCREVER OU PREVER UM ATRIBUTO ESPECIAL CHAMADO CLASSE.**
- **USAMOS CLASSIFICAÇÃO PARA PREVER UMA FRAUDE, DESCOBRIR A QUAL ESPÉCIE UM ANIMAL PERTENCE, PREVER UMA DOENÇA OU CLASSIFICAR UM TIPO DE FUNGO.**



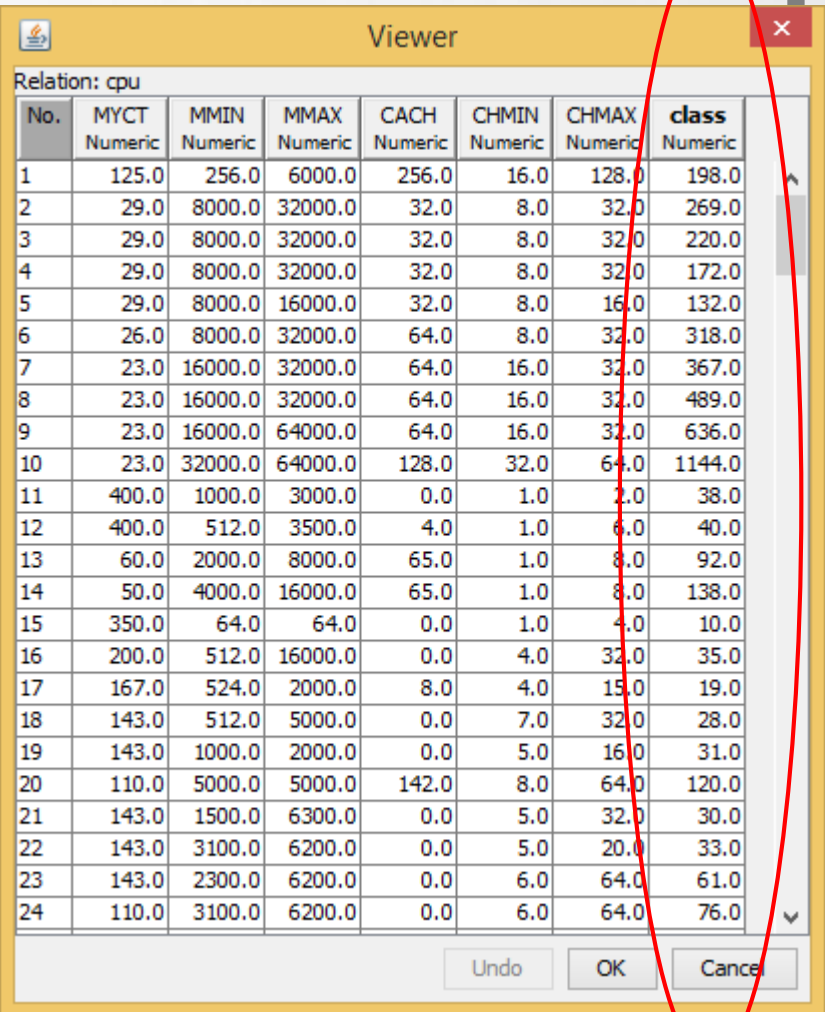
Relation: german\_credit

credits	job	num_dependents	own_telephone	foreign_worker	class
	Nominal	Numeric	Nominal	Nominal	Nominal
2.0	skilled	1.0	yes	yes	good
1.0	skilled	1.0	none	yes	bad
1.0	unskill...	2.0	none	yes	good
1.0	skilled	2.0	none	yes	good
2.0	skilled	2.0	none	yes	bad
1.0	unskill...	2.0	yes	yes	good
1.0	skilled	1.0	none	yes	good
1.0	high q...	1.0	yes	yes	good
1.0	unskill...	1.0	none	yes	good
2.0	high q...	1.0	none	yes	bad
1.0	skilled	1.0	none	yes	bad
1.0	skilled	1.0	none	yes	bad
1.0	skilled	1.0	yes	yes	good
2.0	unskill...	1.0	none	yes	bad
1.0	skilled	1.0	none	yes	good
1.0	unskill...	1.0	none	yes	bad
2.0	skilled	1.0	none	yes	good
3.0	skilled	1.0	none	yes	good
1.0	high q...	1.0	yes	yes	bad
1.0	skilled	2.0	yes	yes	good
3.0	skilled	1.0	yes	yes	good
1.0	skilled	2.0	none	yes	good
2.0	unskill...	2.0	none	no	good

Undo OK Cancel

# REGRESSÃO

- **UM TIPO DE CLASSIFICAÇÃO, PORÉM, ENQUANTO NA CLASSIFICAÇÃO A CLASSE É UM TIPO DE DADO NOMINAL OU CATEGÓRICO, NA REGRESSÃO A CLASSE É NUMÉRICA.**
- **PREVER A ALTURA DE UMA PESSOA A PARTIR DO PESO É UM EXEMPLO DE TAREFA DE REGRESSÃO.**

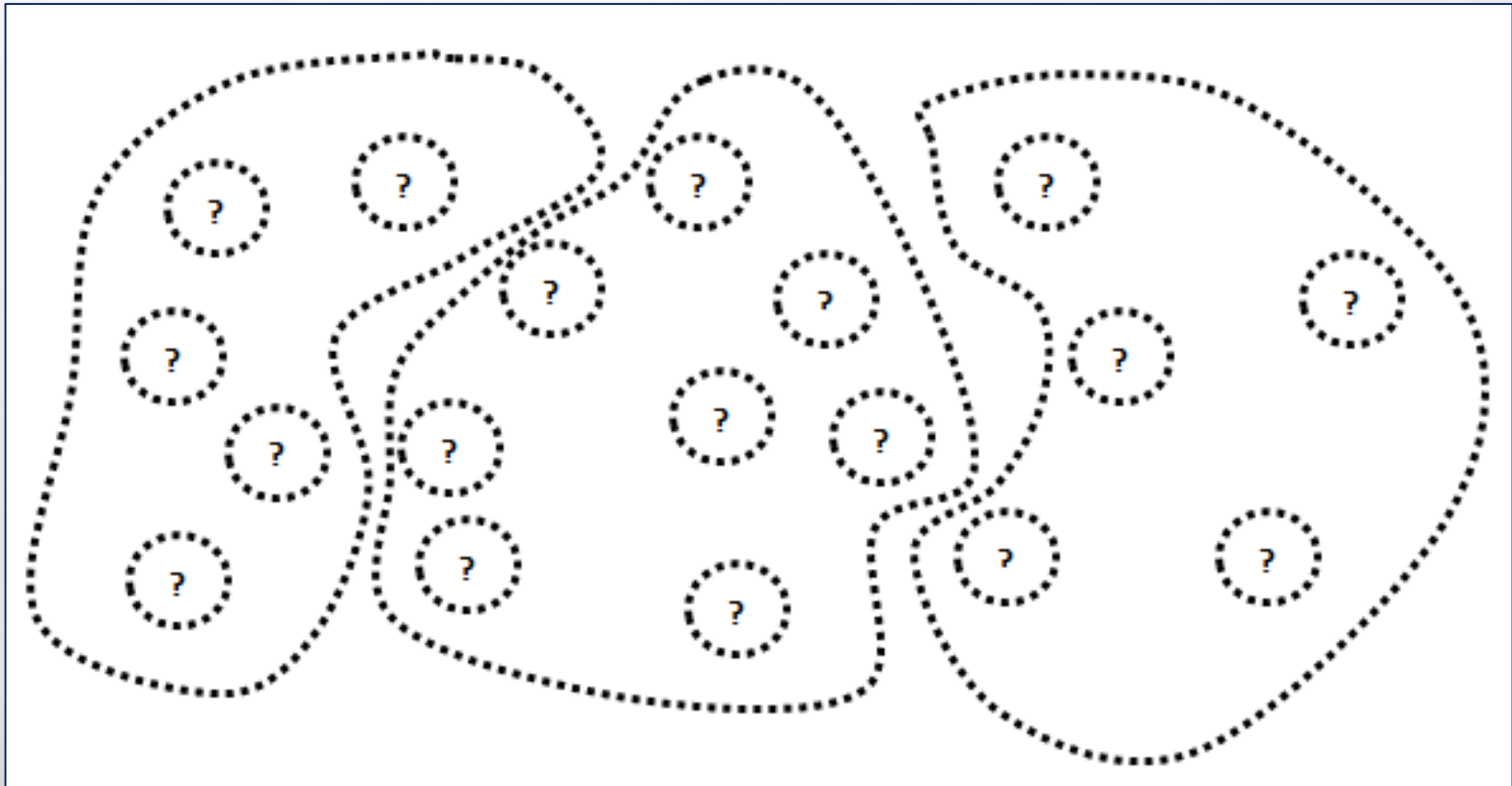


Relation: cpu

No.	MYCT Numeric	MMIN Numeric	MMAX Numeric	CACH Numeric	CHMIN Numeric	CHMAX Numeric	class Numeric
1	125.0	256.0	6000.0	256.0	16.0	128.0	198.0
2	29.0	8000.0	32000.0	32.0	8.0	32.0	269.0
3	29.0	8000.0	32000.0	32.0	8.0	32.0	220.0
4	29.0	8000.0	32000.0	32.0	8.0	32.0	172.0
5	29.0	8000.0	16000.0	32.0	8.0	16.0	132.0
6	26.0	8000.0	32000.0	64.0	8.0	32.0	318.0
7	23.0	16000.0	32000.0	64.0	16.0	32.0	367.0
8	23.0	16000.0	32000.0	64.0	16.0	32.0	489.0
9	23.0	16000.0	64000.0	64.0	16.0	32.0	636.0
10	23.0	32000.0	64000.0	128.0	32.0	64.0	1144.0
11	400.0	1000.0	3000.0	0.0	1.0	2.0	38.0
12	400.0	512.0	3500.0	4.0	1.0	6.0	40.0
13	60.0	2000.0	8000.0	65.0	1.0	8.0	92.0
14	50.0	4000.0	16000.0	65.0	1.0	8.0	138.0
15	350.0	64.0	64.0	0.0	1.0	4.0	10.0
16	200.0	512.0	16000.0	0.0	4.0	32.0	35.0
17	167.0	524.0	2000.0	8.0	4.0	15.0	19.0
18	143.0	512.0	5000.0	0.0	7.0	32.0	28.0
19	143.0	1000.0	2000.0	0.0	5.0	16.0	31.0
20	110.0	5000.0	5000.0	142.0	8.0	64.0	120.0
21	143.0	1500.0	6300.0	0.0	5.0	32.0	30.0
22	143.0	3100.0	6200.0	0.0	5.0	20.0	33.0
23	143.0	2300.0	6200.0	0.0	6.0	64.0	61.0
24	110.0	3100.0	6200.0	0.0	6.0	64.0	76.0

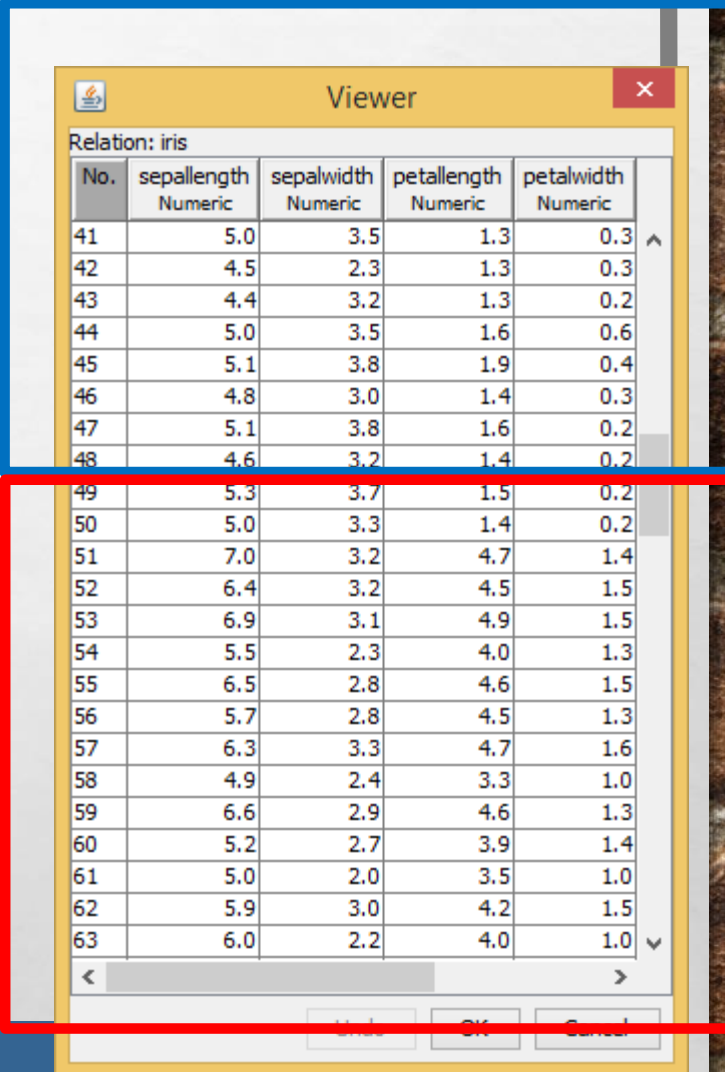
Undo OK Cancel

# AGRUPAMENTOS



# AGRUPAMENTOS

- **NÃO EXISTE CLASSE. O OBJETIVO É CRIAR GRUPOS E ATRIBUIR ÀS INSTANCIAS DOS DADOS NESTES GRUPOS, A PARTIR DAS CARACTERÍSTICAS, OU ATRIBUTOS DESTAS INSTANCIA. EXEMPLOS DE USO: IDENTIFICAR GRUPOS DE CLIENTES E DIRECIONAR CAMPANHAS DE MARKETING ESPECÍFICAS; IDENTIFICAR TENTATIVAS DE ACESSO A REDE; CATEGORIZAR UMA NOVA ESPÉCIE ENTRE OUTROS.**

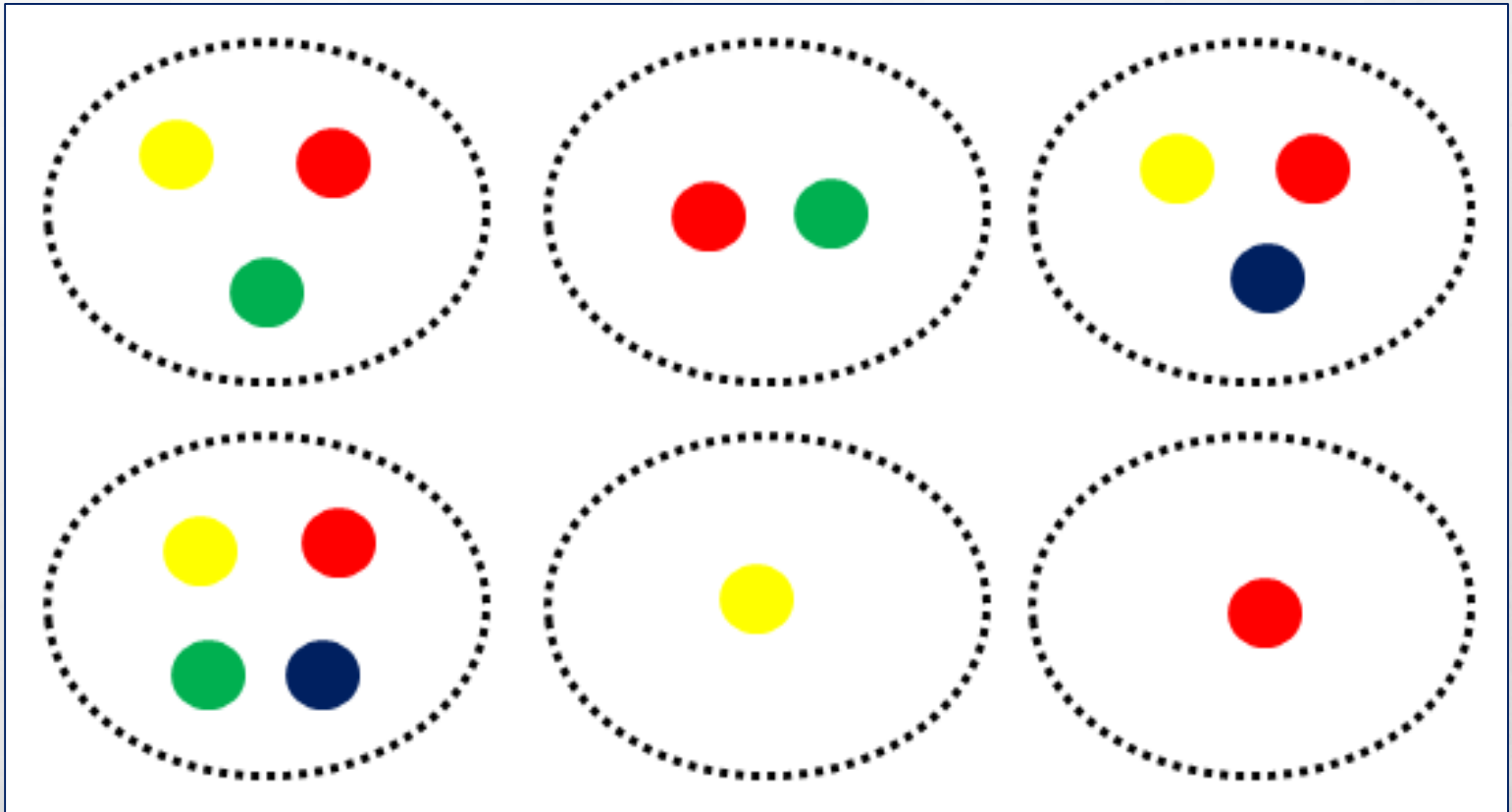


Relation: iris

No.	sepalength Numeric	sepalwidth Numeric	petallength Numeric	petalwidth Numeric
41	5.0	3.5	1.3	0.3
42	4.5	2.3	1.3	0.3
43	4.4	3.2	1.3	0.2
44	5.0	3.5	1.6	0.6
45	5.1	3.8	1.9	0.4
46	4.8	3.0	1.4	0.3
47	5.1	3.8	1.6	0.2
48	4.6	3.2	1.4	0.2
49	5.3	3.7	1.5	0.2
50	5.0	3.3	1.4	0.2
51	7.0	3.2	4.7	1.4
52	6.4	3.2	4.5	1.5
53	6.9	3.1	4.9	1.5
54	5.5	2.3	4.0	1.3
55	6.5	2.8	4.6	1.5
56	5.7	2.8	4.5	1.3
57	6.3	3.3	4.7	1.6
58	4.9	2.4	3.3	1.0
59	6.6	2.9	4.6	1.3
60	5.2	2.7	3.9	1.4
61	5.0	2.0	3.5	1.0
62	5.9	3.0	4.2	1.5
63	6.0	2.2	4.0	1.0



# REGRAS DE ASSOCIAÇÕES



# REGRAS DE ASSOCIAÇÕES

- **BUSCAM A RELAÇÃO ENTRE ITENS. A APLICAÇÃO CLÁSSICA É EM CESTAS DE COMPRAS: QUEM COMPROU PRODUTO A TAMBÉM COMPROU PRODUTO B. PORÉM, TÊM AMPLA APLICAÇÃO EM DIAGNÓSTICOS EM MEDICINA, SENSOS ETC.**

[illegible]



# **SUPERVISIONADO VERSUS NÃO SUPERVISIONADO**

- **AS TAREFAS DE MINERAÇÃO DE DADOS SÃO DITAS SUPERVISIONADAS QUANDO EXISTE UMA CLASSE, OU UM ATRIBUTO ESPECIAL COM O QUAL SE PODE COMPARAR E VALIDAR O RESULTADO.**

# TAREFA NÃO É ALGORITMO

Tarefas	Tipos de Algoritmos	Algoritmos
Classificação	Bayes	NaiveBayes BayesNet
	Rules	OneR Part
Agrupamentos	Por Densidade	DBSCAN
	Centróide	K-means
Regras de Associação	Frequência de Sub- conjuntos	Apriori
	Árvore de Dados	FP Growth