

Conhecendo a base de dados

Transcrição

Agora já temos nosso arquivo "Banco.arff".

Então, reproduzindo o processo, abriremos o "Explorer" e o arquivo do banco em "Weka > Data > Banco.arff".

A primeira informação que veremos serão os atributos que aparecem, referentes às características coletadas de cada cliente, como idade, trabalho, estado civil, educação.

Essas foram 17 informações coletadas para cada cliente, são os nomes das colunas. A última das informações é o depósito, se o cliente fez ou não um depósito na conta. Clicando nessa informação, do lado direito inferior, veremos a quantidade de clientes que efetuou o depósito.

Se colocarmos o mouse sobre a barra azul na tela, veremos a quantidade de "yes", clientes que fizeram o depósito, e em vermelho estarão os clientes que não fizeram o depósito, ou "no". Essa mesma informação aparecerá acima em atributo selecionado ou "Selected attribute".

Do lado esquerdo teremos o nome da nossa base de dados, "Bank", e o número de instâncias. As instâncias serão a quantidade de clientes que coletamos informações, os atributos a quantidade de informações e "Sum of weights" será o mesmo que o número de instâncias.

Com essas informações podemos começar a trabalhar. Poderíamos ter um problema em que em vez de avaliar se o cliente fez ou não um depósito, estudaríamos quanto os clientes depositaram. Nesse caso, teríamos um problema de regressão em que haveria uma tentativa de estimar quanto um cliente depositaria na conta.

Com base nas informações dos clientes, poderíamos tentar agrupá-los. Assim, poderíamos criar uma série de grupos diferentes, com um problema de agrupamento ou *Cluster*.

Nesse caso, temos clientes que fizeram ou não um depósito após serem submetidos a uma campanha de marketing. Então, queremos criar um classificador de clientes, de forma que quando um novo cliente for mostrado, depois de ensinarmos o algoritmo a entender o problema, seja possível classificar se esse cliente vai ou não efetuar o depósito.

Assim, podemos evitar de ligar para um cliente que provavelmente não fará o depósito, ou podemos submetê-lo a uma campanha de marketing diferente, que seja mais favorável a ele.