DM Rule Modeling Challenge – Classify Employees – Corticon – Mike Parish

Challenge Sep-2017

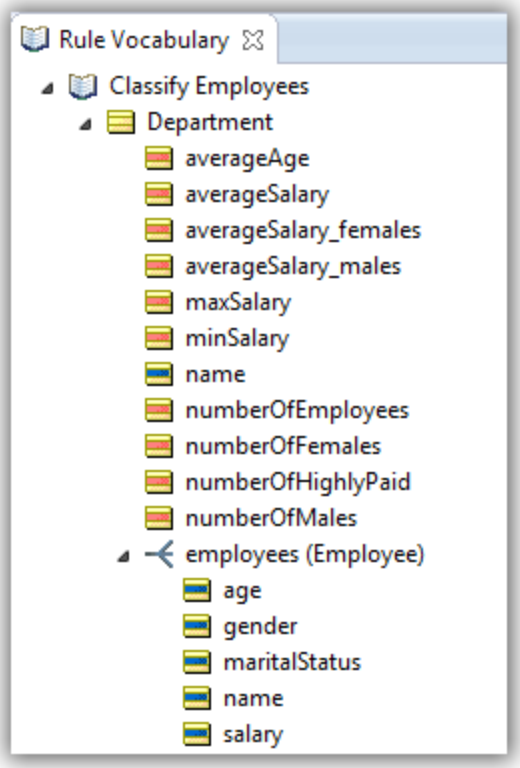
Classify Department Employees

Solutions

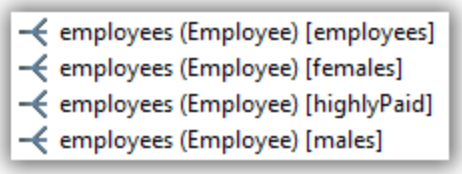
Name	Marital Status	Gender	Age	Salary
Robinson	Married	Female	25	20000
Warner	Married	Male	45	150000
Stevens	Single	Male	24	35000
White	Married	Female	32	75000
Smith	Single	Male	46	110000
Green	Married	Female	28	40000
Brown	Married	Male	32	65000
Klaus	Married	Male	54	85000
Houston	Single	Female	47	35000
Long	Married	Male	29	40000
Short	Single	Male	22	20000
Doe	Single	Female	21	21000

A human resource office has information about all employees in every department including: salary, marital status, age, etc. Help the office to create a decision model that for each department calculates minimal, maximal, and average salaries along with a number of high-paid employees using rules like "Salary > 85000".

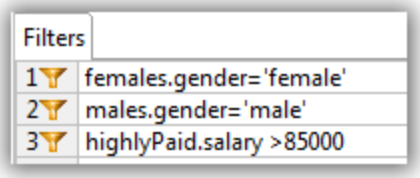
We can define a data model something like this where there is a one to many association between departments and their employees. Blue attributes are data passed to the rules. Orange attributes are those calculated by the rules.



In Corticon we can refer to different subsets of the employees by using aliases:

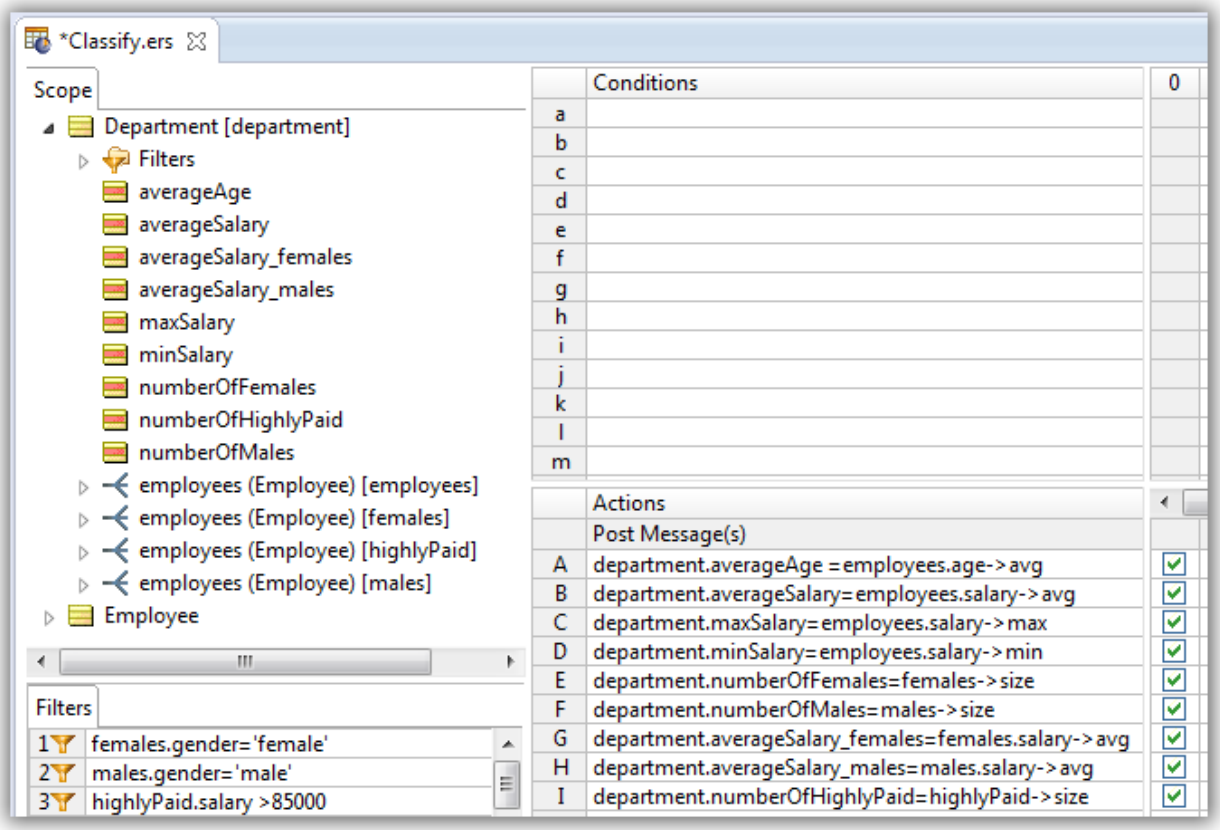


Membership of these subsets is defined by filters:



When no filter is defined the alias refers to the entire collection.

The Rule Sheet



We can easily create more classifications by adding aliases to the employees and defining the appropriate filter.

Then we can refer to those aliases to perform various calculations using the operators

->avg, ->max, ->min, ->sum

Even more powerful are the operators that can be applied to sorted collections of data:

->at, ->first, ->first(n), ->last, ->last(n), ->subsequence(from,to), ->trend

Corticon can then easily support the following kinds of classification questions

Conditions
Is there any person over 65?
Are there any persons between 13 and 19?
Are there more than 2 persons between 20 and 65?
Are there any males under 13?
Is the average age of the oldest 3 males between 21 and 45 greater than 25?
Is the sum of incomes for people over 65 greater than 10
Is the lowest income for people over 65 greater than zero?
Is the highest income for males over 65 greater than 10?
Is the sum of the incomes for the youngest 5 males greater than zero?
Is the youngest person over 65 a male?
Are all the people over 65 male?
Are there any people over 65 that are not male?
Are there any people over 65 that are male?
Is the highest income of the youngest 5 males greater than 10000?
Is the person who has the lowest income over a threshold a male

Conditions
person[age>65]->notEmpty
person[age in 13..19]->notEmpty
person[age in 20..65]->size > 2
person[age <13, gender='male']->notEmpty
person.income[age in 21..45, gender='male'] ->sortedBy(age) ->first(3) ->avg >25
person.income[age>65]->sum >10
person.income[age>65]->min >0
person.income[age>65,gender='male'] ->max >10
person.income[gender='male'] ->sortedBy(age) ->first(5) ->sum > 0
person.gender[age>65] ->sortedBy(age)->first = 'male'
person.gender[age>65] ->allContain('male')
person[age>65,gender<>'male'] ->notEmpty
person[age>65,gender='male'] ->notEmpty
person.income[gender='male'] ->sortedBy(age) ->first(5) ->sortedByDesc(income) ->first > 10000
person.gender[income >Parameter.value] ->sortedBy(income)->first = 'male'