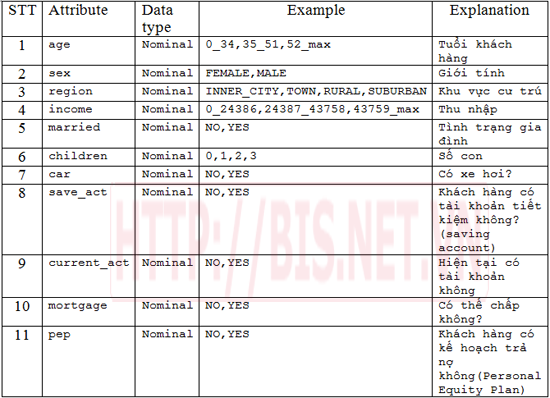
**Khai phá luật kết hợp với Weka**

(Association Rule Mining with WEKA )

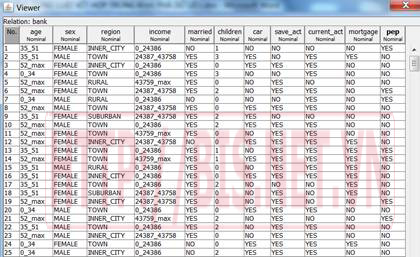
Trong lĩnh vực Data Mining, mục đích của luật kết hợp (Association Rule - AR) là tìm ra các mối kết hợp (Association) hay tương quan (Correlation)  giữa các đối tượng trong khối lượng lớn dữ liệu. Ứng dụng của luật kết hợp rất phổ biến trong nhiều lĩnh vực, nhất là trong kinh doanh như Market Basket Analysis (Cross selling**,**Product placement, Affinity promotion, Customer behavior Analysis).

Bài viết này minh họa việc sử dụng thuật toán Apriori để phát hiện các luật kết hợp trên phần mềm Weka.

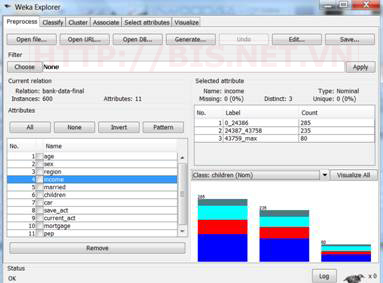
Dữ liệu được dùng trong ví dụ này là dữ liệu về khách hàng ngân hàng (bank.arff) gồm 11 thuộc tính và 600 khách hàng (samples, instances, patterns). Cấu trúc file bank.arff như sau:



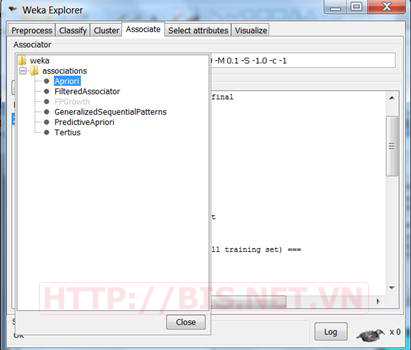
*Dữ liệu về một số khách hàng trong file bank.arff như sau*



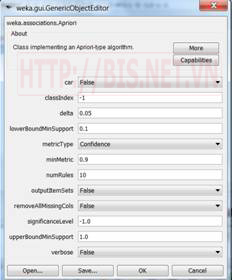
***Nạp dữ liệu (nạp file bank.arff)***



***Chọn tab Association và chọn thuật toán Apriori***



**Thiết lập các tham số (numRules, Support, Confidence,…)**



***Giải thích một số tham số chính của thuật toán Apriori sinh luật kết hợp***

*lowerBoundMinSupport:* Cận dưới của minimum support.

*metricType:* Có 4 loại metricType là Confidence, Lift , Leverage và Conviction.

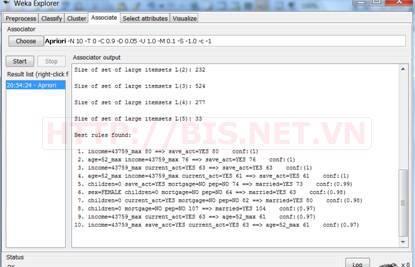
*Minimum metric score*: Chỉ quan quan đến các luật có metric score cao hơn giá trị này.

*numRules:* Số luật muốn tìm (các luật sẽ được sắp xếp theo thứ tự giảm dần của metric score.

*significanceLevel:* Mức ý nghĩa (chỉ dùng khi metricType là confidence).

upperBoundMinSupport: Cận trên của minimum support (bắt đầu lặp lại việc giảm minimum support từ upperBoundMinSupport đến lowerBoundMinSupport).

***Các luật kết hợp thu được***



*10 luật tốt nhất được sinh ra bởi thuật toán Apriori trên dữ liệu bank.arff như sau (sắp xếp giảm dần theo confidence):*

1*. income=43759\_max 80 ==> save\_act=YES 80    conf:(1)*

*2. age=52\_max income=43759\_max 76 ==> save\_act=YES 76    conf:(1)*

*3. income=43759\_max current\_act=YES 63 ==> save\_act=YES 63    conf:(1)*

*4. age=52\_max income=43759\_max current\_act=YES 61 ==> save\_act=YES 61    conf:(1)*

*5. children=0 save\_act=YES mortgage=NO pep=NO 74 ==> married=YES 73    conf:(0.99)*

*6. sex=FEMALE children=0 mortgage=NO pep=NO 64 ==> married=YES 63    conf:(0.98)*

*7. children=0 current\_act=YES mortgage=NO pep=NO 82 ==> married=YES 80    conf:(0.98)*

*8. children=0 mortgage=NO pep=NO 107 ==> married=YES 104    conf:(0.97)*

*9. income=43759\_max current\_act=YES 63 ==> age=52\_max 61    conf:(0.97)*

*10. income=43759\_max save\_act=YES current\_act=YES 63 ==> age=52\_max 61    conf:(0.97)*

*Luât kết hợp số 1 cho biết rằng nếu thu nhập (income) lớn hơn hoặc bằng 43759 thì khách hàng có tài khoản tiết kiệm (save\_act =YES) với độ tin cậy là 1.*