**A – Báo cáo viết**

**Yêu cầu 1 – Khảo sát luật kết hợp trên dữ liệu vote.arff:**

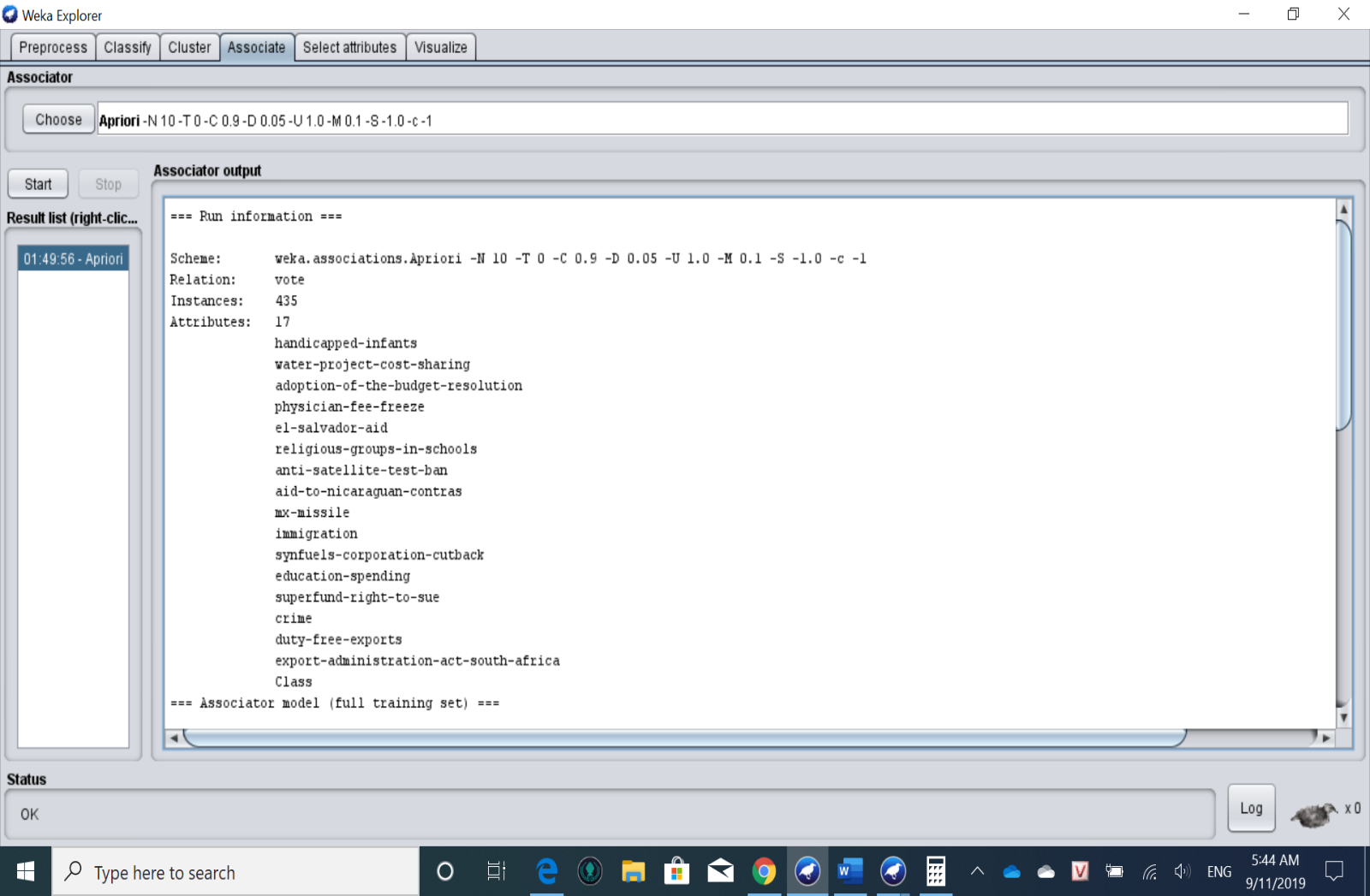
## **Chạy giải thuật Apriori với tham số mặc định. 10 dòng cuối cùng của phần kết quả, bên dưới dòng “Best rules found”, là 10 luật được chọn hiển thị trong số các luật đã phát sinh. Độ tin cậy (confidence) của luật 10 là 0.96. Giá trị này được tính như thế nào? Hãy trình bày công thức và thế giá trị cụ thể vào các phép tính.**

Công thức tính độ tin cậy của luật X Y: confidence = Pr(Y ) =

Theo hình trên ta có tổng số mẫu tiền đề là 258, số mẫu dự đoán đúng là: 246

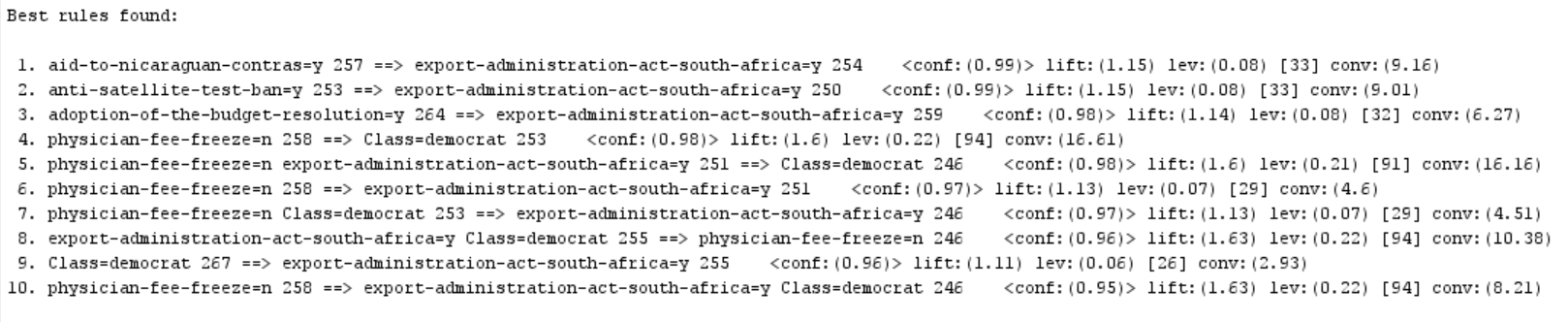
Confidence của **luật 10** thay số vào công thức trên ta có: conf = = 0.95

## **Có bao nhiêu mẫu được xét khi cần tính độ hỗ trợ (support) của luật 8 ?**



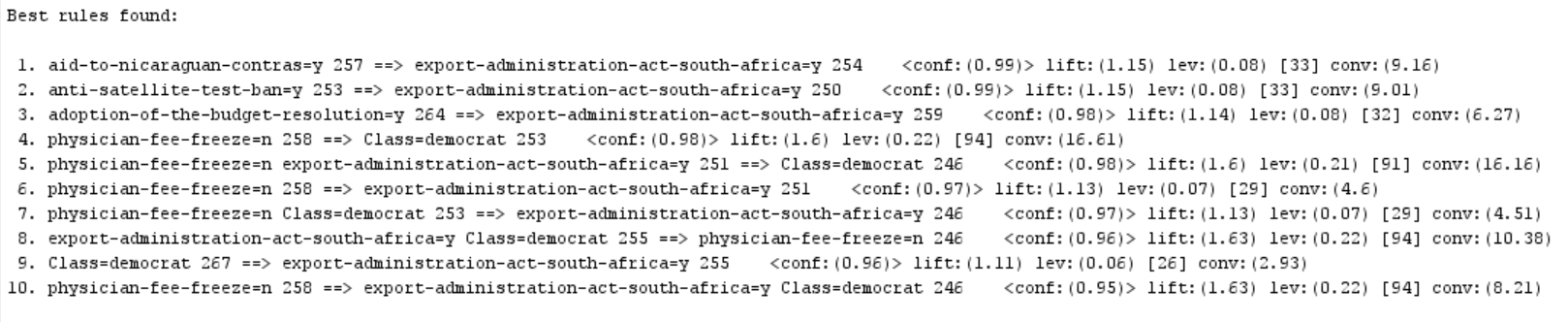
* Có 435 mẫu

## **“Số lượng mẫu có thể áp dụng luật” nghĩa là gì? Giải thích bằng ví dụ là luật 7.**



* Số lượng mẫu áp dụng luật là số lượng mẫu tiền đề trong giao dịch, ta dùng để xét các luật kết hợp
* Giải thích:
* **Số lượng mẫu có thể áp dụng luật:** là số mẫu đưa vào tiền đề để xét luật kết hợp
* **Giải thích**: Như hình trên ở luật 7, 253 mẫu **physician-fee-freeze=n Class=democrat** là mẫu tiền đề, sau khi chạy thuật toán Apriori thì ứng với 253 mẫu này sẽ kết hợp được với mẫu **export-administration-act-south-africa=y** và tổng các mẫu cho luật kết hợp là 246 (tiền đề và hệ quả).

## **“Số lượng mẫu được dự đoán chính xác bởi luật” có nghĩa là gì? Giải thích thông qua ví dụ là luật 9.**



* **Số lượng mẫu được dự đoán chính xác bởi luật:** là số lượng mẫu thõa tiền đề và hệ quả của luật kết hợp.
* **Giải thích:** ở hình trên luật 9, thuộc tính **Class=democrat 267(1) ==> export-administration-act-south-africa=y 255(2), 255** là số lượng mẫu thỏa (1) và (2)

## **Khảo sát phần mô tả tham số của Apriori bằng cách nhấn vào nút “More” trong cửa sổ tùy chọn tham số GenericObjectEditor. Hãy thử thay đổi số luật được hiển thị trong kết quả. Bạn có nghĩ rằng số luật được phát sinh có thể nhiều hơn 100 không? Giải thích. Chụp màn hình các tình huống số lượng luật hiển thị khác nhau làm minh chứng.**

* Thử thay đổi số luật được hiển thị trong kết quả. Như trong hình lấy ví dụ là 100 luật



* Số lượng luật phát sinh có thể nhiều hơn 100 bởi vì: min
* Hình ảnh số lượng luật hiển thị khác nhau: trong hình đã thử phát sinh ra 1000 luật

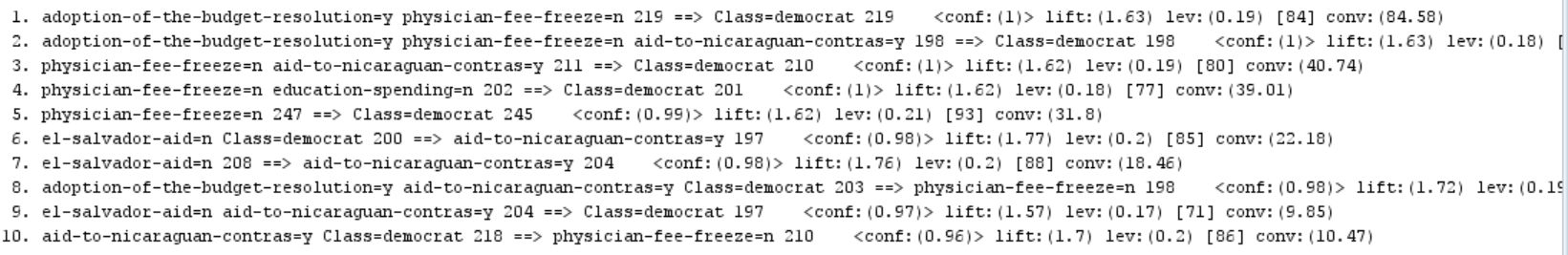


## **“Luật tốt nhất” nghĩa là gì? Tiêu chí nào được dùng để xác định các luật tốt nhất?**

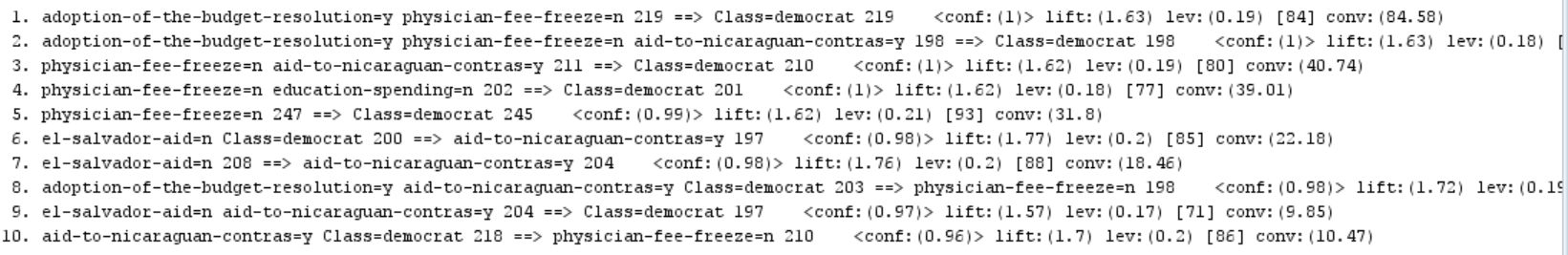
* **Luật tốt nhất** : khi một tiền đề nào đó xuất hiện sẽ dẫn đến một hệ quả nào đó cũng xuất hiện theo, mà số mẫu được luật dự đoán chính xác nhiều nhất so với những luật khác có cùng tiền đề.
* **Tiêu chí để xác định các luật tốt nhất :** độ confidence.

## **Luật nào nói về khả năng một Hạ nghị sĩ không bỏ phiếu cho ‘el-salvador-aid’ thì người này sẽ bỏ phiếu cho ‘aid-to-nicaraguan-contras’? Tương tự, luật nào có kèm theo điều kiện về đảng phái của Hạ nghị sĩ này?**

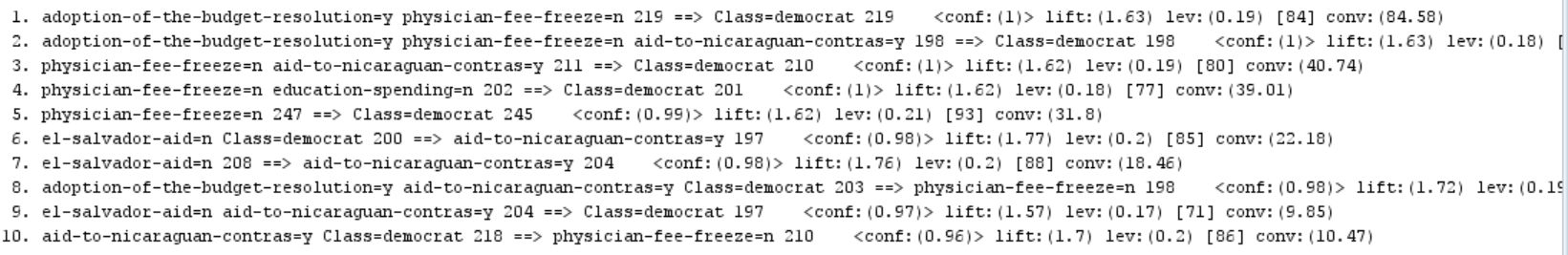
* Luật 7 thõa yêu cầu: **một Hạ nghị sĩ không bỏ phiếu cho ‘el-salvador-aid’ thì người này sẽ bỏ phiếu cho ‘aid-to-nicaraguan-contras’(**xem ảnh ở dưới)



* Luật nói về khả năng một Hạ nghị sĩ **không** bỏ phiếu cho **‘el-salvador-aid’** thìngười này sẽ **bỏ phiếu** cho **‘aid-to-nicaraguan-contras’** và kèm theo **điều kiện đảng phải** đó là **luật số 6,** đang phái cho Hạ nghị sĩ: **democrat**



## **Một số luật trong 10 luật tốt nhất có vế phải là ‘Class = democrat’. Vấn đề này nói lên điều gì về thói quen bỏ phiếu của các Hạ nghị sĩ đảng Dân chủ?**



* Vấn đề nói lên

**Yêu cầu 2 – Khảo sát luật kết hợp trên dữ liệu weather.nominal.arff:**

**Yêu cầu 3 – Khảo sát luật kết hợp trên dữ liệu supermarket.arff:**

* Dựa vào hình ở trên thì suy ra được thói quen của các Hạ sĩ đảng dân chủ như sau:
  + **Bỏ phiếu** cho **adoption-of-the-budget-resolution** và không bỏ phiếu cho **Physician-fee-freeze** là **Hạ sĩ của đảng dân chủ.**
  + **Bỏ phiếu** cho **adoption-of-the-budget-resolution, không bỏ phiếu** cho **Physician-fee-freeze** và **bỏ phiếu** cho **aid-to-nicaraguan-contras** là **Hạ sĩ của đảng dân chủ.**
  + **Không bỏ phiếu** cho **physician-fee-freeze** và **bỏ phiếu**   
    **aid-to-nicaraguan-contras** là **Hạ sĩ của đảng dân chủ.**
  + **Không bỏ phiếu** cho **physician-fee-freeze** và **không bỏ phiếu** cho   
    **education-spending** là **Hạ sĩ của đảng dân chủ.**
  + **Không bỏ phiếu** cho **physician-fee-freeze** là **Hạ sĩ của đảng dân chủ.**
  + **Không bỏ phiếu** cho **el-salvador-aid** và **bỏ phiếu** cho  
    **aid-to-nicaraguan-contras** là **Hạ sĩ của đảng dân chủ.**

**Yêu cầu 2 – Khảo sát luật kết hợp trên dữ liệu vote.arff:**

**2.1. Xét luật temperature=hot ==> humidity=normal. Số lượng mẫu áp dụng luật là bao nhiêu? Số lượng mẫu thỏa mãn luật là bao nhiêu? Tính độ hỗ trợ và độ tin cậy của luật.**

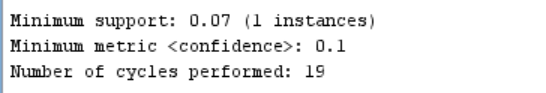
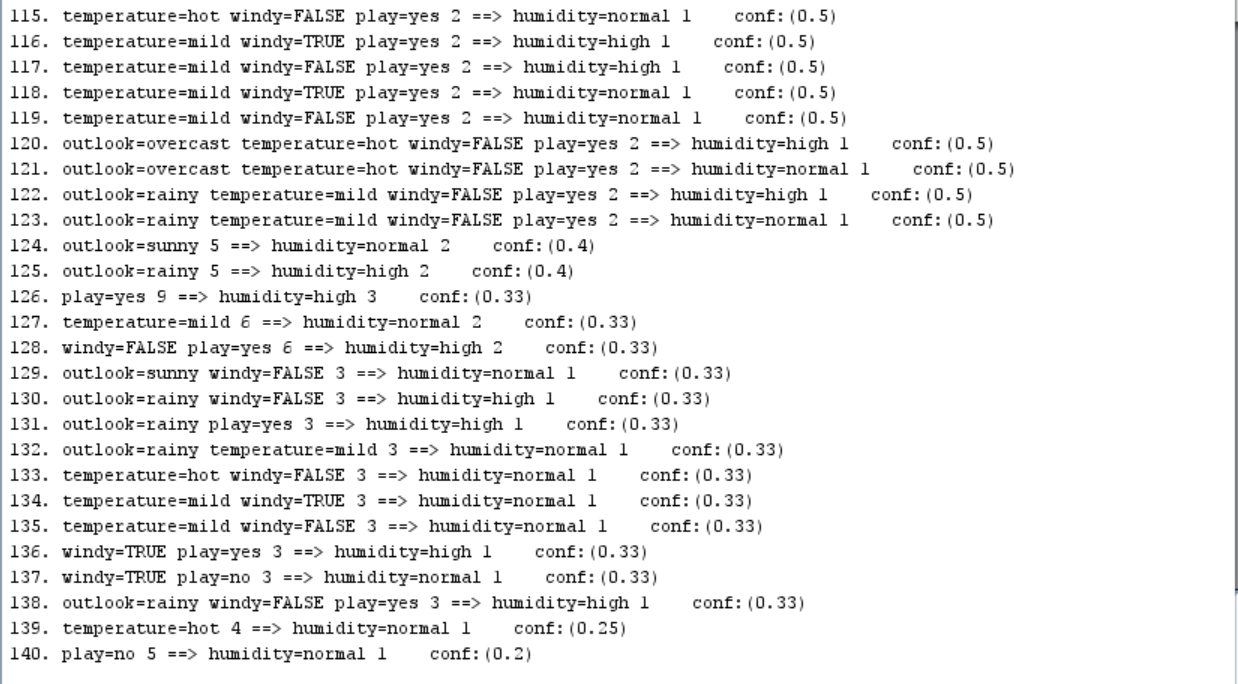
* Số lượng mẫu áp dụng luật là: 4
* Số lượng mẫu thỏa mãn luật là: 1
* Độ hỗ trợ của luật: sup = Pr(Y ) = = = 0.07
* Độ tin cậy của luật: conf = Pr(Y ) = = = 0.25

**2.2. Xét luật temperature=hot humidity=high ==> windy=TRUE. Số lượng mẫu áp dụng luật là bao nhiêu? Số lượng mẫu thỏa mãn luật là bao nhiêu? Tính độ hỗ trợ và độ tin cậy của luật.**

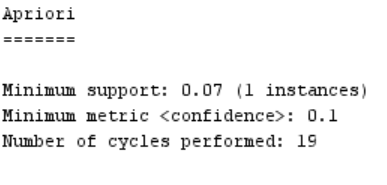
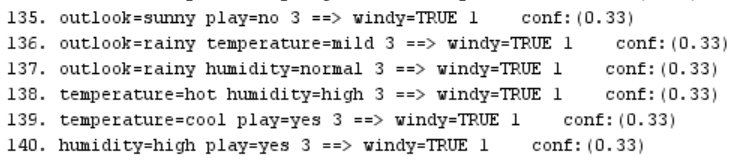
* Số lượng mẫu áp dụng luật là: 3
* Số lượng mẫu thỏa mãn luật là: 1
* Độ hỗ trợ của luật: sup = Pr(Y ) = = = 0.07
* Độ tin cậy của luật: conf = Pr(Y ) = = = 0.3

**2.3. Điều chỉnh số lượng luật được hiển thị, giá trị minsup, và giá trị minconf, nếu cần thiết. Luật ở câu 2.1 và câu 2.2 nằm ở vị trí thứ mấy trong danh sách luật được tìm thấy? Chụp màn hình có hiển thị phần luật tương ứng làm minh chứng.**

* Luật của câu 2.1: nằm ở **dòng 139** (có đánh dấu trong hình)



* Luật của câu 2.2 nằm **ở dòng 138** (có đánh dấu trong hình)

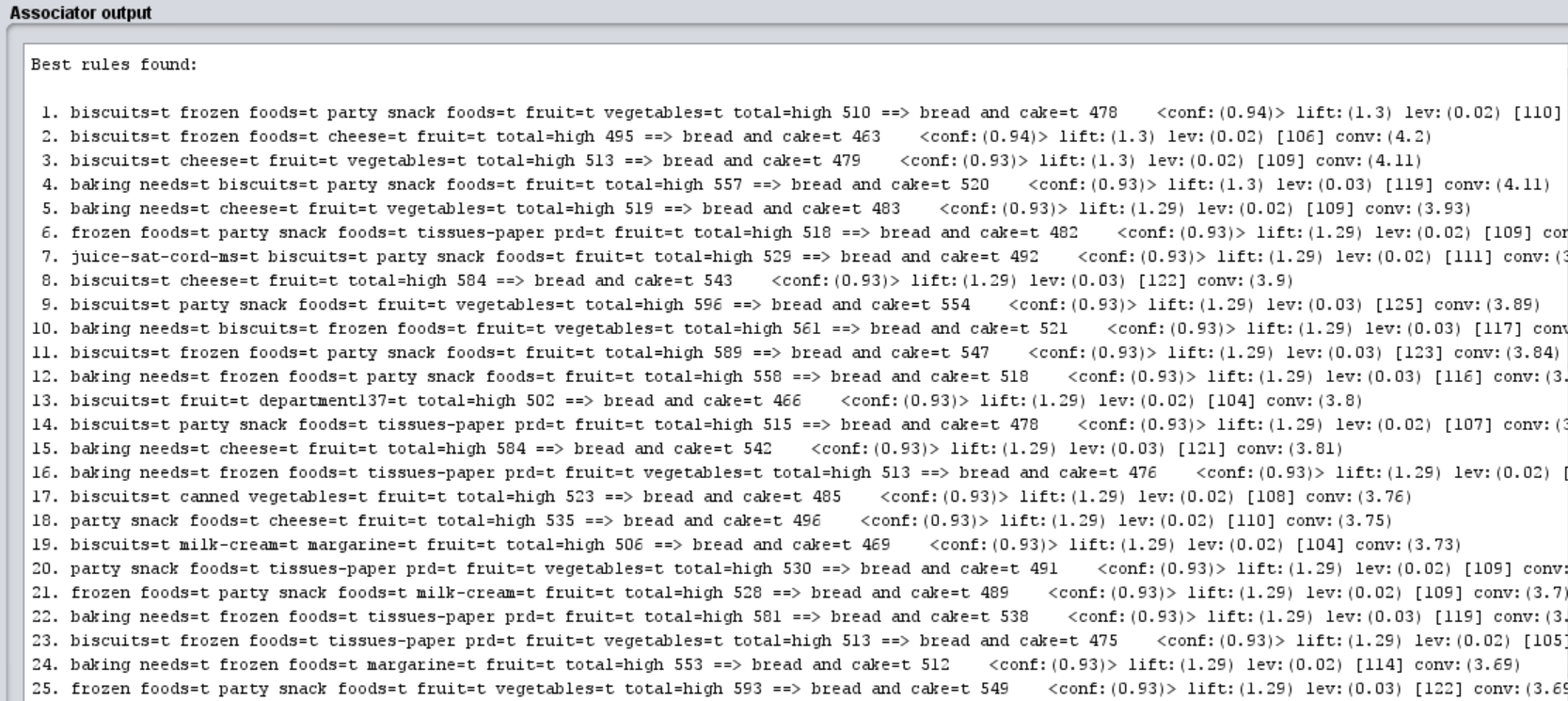


**2.4. Một luật có thể có hai (hay nhiều) thuộc tính ở vế phải được không? Ví dụ, outlook=sunny temperature=cool ==> humidity=normal play=yes Hãy đưa ra 5 luật khác luật đã nêu ở trên để làm ví dụ**

* Một luật có thể có hai (hay nhiều) thuộc tính ở vế phải, ví dụ:
  + temperature=hot windy=TRUE ==> outlook=sunny play=no
  + temperature=cool windy=FALSE ==> humidity=normal play=yes
  + windy=FALSE play=no ==> outlook=sunny humidity=high
  + outlook=sunny temperature=hot 2 ==> humidity=high play=no
  + temperature=hot play=no 2 ==> outlook=sunny humidity=high

**Yêu cầu 3 – Khảo sát luật kết hợp trên dữ liệu supermarket.arff**

Phát sinh ra 30 luật kết hợp



**3.1. Khảo sát một vài luật được phát sinh và mô tả một quan sát về thói quen mua sắm của khách hàng siêu thị mà bạn rút ra được từ việc khảo sát này. Dẫn chứng những luật kết hợp liên quan đến quan sát được đưa ra.**

**-** Khảo sát 5 luật trong tổng 30 được phát sinh:

* **Luật 1**:

**biscuits=t frozen foods=t party snack foods=t fruit=t vegetables=t total= hight 510 break and cake = t 478 <conf =(0.94)>**  
Khách hàng mua **biscuits**, **frozen foods**, **party snack foods fruit**, **vegetables** và   
**total= hight** thì sẽ mua **break and cake.**

* **Luật 2**:

**Biscuits=t frozen foods=t cheese=t fruit=t total=hight 495 break and cake=t 463 <conf=(0.94)>**

Khách hàng mua **biscuits**, **frozen foods**, **cheese**, **fruit** và **total= hight** thì sẽ mua **break and cake.**

* **Luật 7**:

**juice-sat-cord-ms=t biscuits=t party snack foods=t fruit=t total=hight 529**

**bread and cake=t 492 <conf=(0.93)>**

Khách hàng **mua juice-sat-cord-ns**, **biscuits**, **party snack foods**, **fruit**, **total = hight** thì sẽ mua **break and cake.**

* **Luật 16:**

**Baking needs=t frozen foods=t tissues-paper prd=t fruit=t vegetable=t total = hight 513 bread and cake=t 476 <conf=(0.93)>**

Khách hàng mua **baking needs**, **frozen foods**, **tissues-paper prd**, **fruit** và **total = hight** thì sẽ mua **bread and cake.**

* **Luật 29:**

**camed vegetables=t frozen foods=t fruit=t total=hight 521 bread and cake=t 482 <conf=(0.93)>**

Khách hàng mua **canned vegetables**, **frozen foods, fruit** và **total = hight** thì sẽ mua **bread and cake**

* Như chúng ta có thể thấy 5 luật có điểm chung đó là **total=hight** cho thấy khi khách hàng mua nhiều thì sẽ mua thêm **bread and cake**.

**3.2. Tương tự câu hỏi trên, hãy mô tả quan sát thứ hai mà bạn có được, đồng thời dẫn chứng những luật kết hợp liên quan.**

**-** Như trong hình (ở đầu yêu cầu 3) chúng ta thấy mọi luật ở tiền đề khác nhau đều cho ra hệ quả đó là **bread and cake**

**-** Một số luật kết hợp dẫn chứng:

* **Luật 3:**

**biscuits=t cheese=t fruit=t vegetables=t total=hight 513 break and cake=t 463 <conf=(0.93)>**

* **Luật 10:**

**baking needs=t biscuits=t frozen foods=t fruit=t vegetable=t total=hight 561 break and cake 521 <conf=(0.93)>**

* **Luật 18:**

**party snack foods=t cheese=t fruit total=hight 535 break and cake 496 <conf=(0.93)>**

* **Luật 23:**

**biscuits=t frozen foods=t tissues-paper prd=t vegetabes=t total=hight 513 break and cake 475 <conf=(0.93)>**

* **Luật 30:**

**biscuits=t milk-cream=t margarine=t vegetables=t total=hight 507 break and cake 496 <conf=(0.93)>**

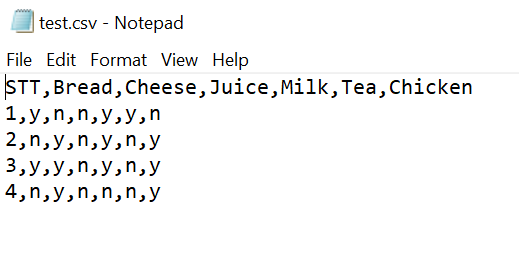
**3.3. Các quan sát của bạn ở hai câu hỏi trên có gợi ý gì cho nhà quản lý siêu thị nhằm đưa ra hành động có ích cho việc kinh doanh? Nếu có, giải thích quan sát hữu ích như thế nào đến hoạt động của siêu thị. Nếu không, giải thích tại sao quan sát không có giá trị.**

**-** Hoạt động có ích cho việc kinh doanh:

* Cần sắp xếp các mặt hàng ở tiền đề và hệ quả theo luật gần nhau để tăng lượng người mua sản phẩm đồng thời tăng doanh thu
* Hoặc có thể giảm giá cho **bread and cake** khi mua những mặc hàng ở tiền và mua thêm bread and cake

**B – Nội dung thực hiện cài đặt:**

* Cách thực thi chương trình: vì nhóm em có dùng đến thư viện **numpy** nên khi thực thi chương trình cô chịu khó cài đặt vào máy.
* Chương trình thực thi với **WindowsPowerShell**
* **Hình minh họa**

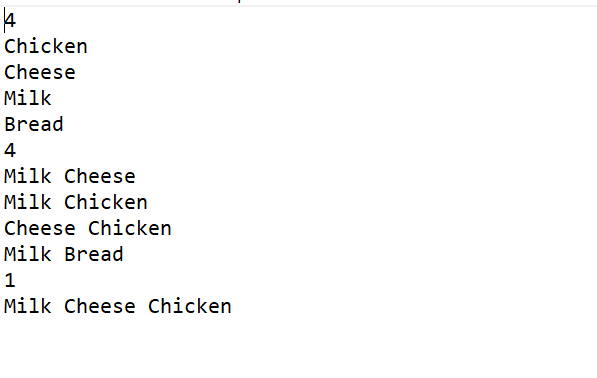
**file text**

**Thực thi chương trình:**

**A screenshot of a cell phone

Description automatically generated**

**File FI.txt**



**File AR.txt** A screenshot of a cell phone

Description automatically generated