THỰC HÀNH CÂU 3

CSDL về Nhân viên được cho trong bảng sau:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Giới tính (GT) | Tuổi (T) | Năng lực làm việc (NL) | Đã lập gia đình (LGD) | Thu nhập (TN) | Thăng chức (TC) |
| 1 |  | Nữ | 20..25 | Giỏi | Rồi | Rất cao | Có |
| 2 |  | Nam | 20..25 | Khá | Chưa | Khá | Không |
| 3 |  | Nữ | 26..30 | Giỏi | Chưa | Khá | Có |
| 4 |  | Nữ | 31..40 | T.Bình | Chưa | T.Bình | Có |
| 5 |  | Nam | 26..30 | T.Bình | Rồi | Rất cao | Không |
| 6 |  | Nữ | 26..30 | Khá | Chưa | Cao | Không |
| 7 |  | Nữ | 31..40 | Khá | Chưa | T.Bình | Không |
| 8 |  | Nam | 26..30 | Khá | Rồi | Cao | Có |
| 9 |  | Nữ | >40 | Giỏi | Rồi | T.Bình | Không |
| 10 |  | Nữ | 26..30 | Giỏi | Chưa | Khá | Có |

Cho B= {Tuổi, Năng lực làm việc, Thăng chức}. Hãy tìm tất cả các luật kết hợp có về phải chỉ gồm thuộc tính Thăng chức (TC) thỏa ngưỡng minsup = 30% và minconf = 80%.

Vì B = {Tuổi, Năng lực làm việc, Thăng chức} nên ta có bảng dữ liệu mới là:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Tuổi (T) | Năng lực làm việc (NL) | Thăng chức (TC) |
| 1 | 20..25 | Giỏi | Có |
| 2 | 20..25 | Khá | Không |
| 3 | 26..30 | Giỏi | Có |
| 4 | 31..40 | T.Bình | Có |
| 5 | 26..30 | T.Bình | Không |
| 6 | 26..30 | Khá | Không |
| 7 | 31..40 | Khá | Không |
| 8 | 26..30 | Khá | Có |
| 9 | >40 | Giỏi | Không |
| 10 | 26..30 | Giỏi | Có |

Đưa dữ liệu về dạng giống như câu 2:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| STT | Tuổi (T) | | | | Năng lực làm việc (NL) | | | Thăng chức (TC) | |
| 20..25 | 26..30 | 31..40 | >40 | Giỏi | Khá | Trung bình | Có | Không |
| 1 | x |  |  |  | x |  |  | x |  |
| 2 | x |  |  |  |  | x |  |  | x |
| 3 |  | x |  |  | x |  |  | x |  |
| 4 |  |  | x |  |  |  | x | x |  |
| 5 |  | x |  |  |  |  | x |  | x |
| 6 |  | x |  |  |  | x |  |  | x |
| 7 |  |  | x |  |  | x |  |  | x |
| 8 |  | x |  |  |  | x |  | x |  |
| 9 |  |  |  | x | x |  |  |  | x |
| 10 |  | x |  |  | x |  |  | x |  |

1. **Tìm tất cả các tập phổ biến từ mẫu dữ liệu trên bằng thuật toán Apriori**

* Với min\_sup = 30% và tổng số giao dịch là 10 => Tần số xuất hiện tối thiểu của phần tử để thỏa min\_sup là 3.
* Tập các ứng viên 1 phần tử và tần số xuất hiện của nó (support count) tương ứng như sau:
* C1 = {{20..25}:2, {26..30}:5, {31..40}:2, {>40}: 1, Giỏi: 4, Khá: 4, Trung bình: 2, Có: 5, Không: 5}
  + - * Các tập phổ biến thỏa min\_sup 3 là: {26..30}, Giỏi, Khá, Có, Không
      * Các tập phổ biến 1 phần tử (tập các ứng viên thoải min\_support\_count):
* L1 = {{26..30}, {Giỏi}, {Khá}, {Có}, {Không}}
* Tập các ứng viên 2 phần tử và tần số xuất hiện của nó
* C2 = L1⋈L1 = {{26..30, Giỏi}: 2, {26..30, Khá}: 2, {26..30, Có}:3, {26..30, Không}: 2, {Giỏi, Có}:3, {Giỏi, Không}:1, {Khá, Có}:1, {Khá, Không}:3}
  + - * Các tập phổ biến thỏa min\_sup 3 là: {26..30, Có}, {Giỏi, Có}, {Khá, Không}
      * Các tập phổ biến 2 phần tử (tập các ứng viên thoải min\_support\_count):
* L2 = {{26..30, Có}, {Giỏi, Có}, {Khá, Không} }
* Tập các ứng viên 3 phần tử và tần số xuất hiện của nó
* C3 = L2⋈L2 = {{26..30, Giỏi, Có}: 2}
  + - * Các tập phổ biến thỏa min\_sup 3 là: không có
      * Các tập phổ biến 3 phần tử (tập các ứng viên thoải min\_support\_count):
* L3 =
* Vì L3 không có tập hợp nào nên không tiến hành tìm tập ứng cử viên 4 phần tử.

1. **Tìm luật có về phải là thăng chức**

* Xét tập phổ biến {26..30, Có} có min\_conf = 80%
* {26..30}: 5, {Có}: 5, {26..30, Có}: 3
* Ta có

|  |  |
| --- | --- |
| 26..30🡪Có | 3/5= 60% |

* Với min\_conf = 80% dựa vào bảng trên ta không có luật kết hợp thỏa yêu cầu min\_conf. Nên tập {26..30, Có} không có luật kết hợp có về phải là thuộc tính Thăng chức.
* Xét tập phổ biến {Giỏi, Có} có min\_conf = 80%
* {Giỏi}: 4, {Có}: 5, {Giỏi, Có}: 3
* Ta có

|  |  |
| --- | --- |
| Giỏi🡪Có | 3/4= 75% |

* Với min\_conf = 80% dựa vào bảng trên ta không có luật kết hợp thỏa yêu cầu min\_conf. Nên tập {Giỏi, Có} không có luật kết hợp có về phải là thuộc tính Thăng chức.
* Xét tập phổ biến {Khá, Không} Không min\_conf = 80%
* {Khá}: 4, {Không}: 5, {Khá, Không}: 3
* Ta có:

|  |  |
| --- | --- |
| Khá🡪Không | 3/4= 75% |

* Với min\_conf = 80% dựa vào bảng trên ta không có luật kết hợp thỏa yêu cầu min\_conf. Nên tập {Khá, Không} không có luật kết hợp có về phải là thuộc tính Thăng chức.
  + - * Không có luật kết hợp cần tìm với min\_conf = 80%.