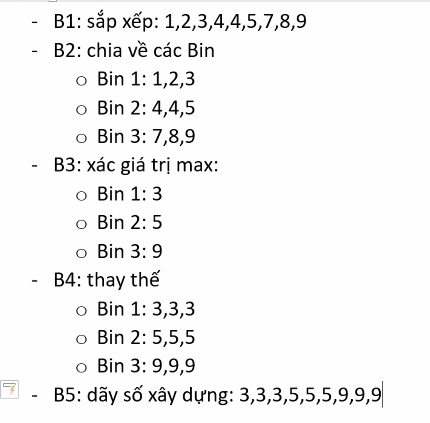
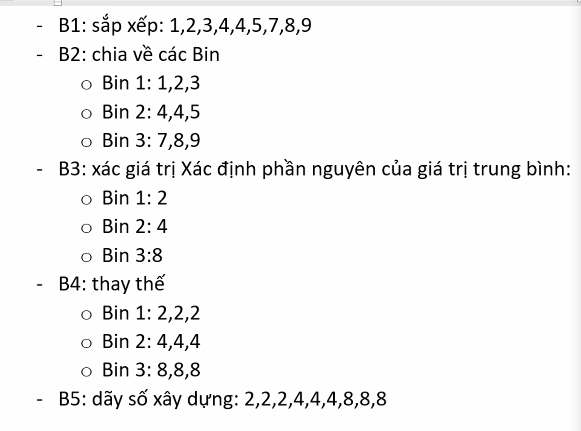
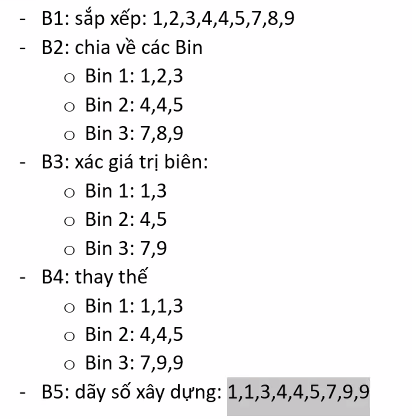
B1: giảm nhiễu dữ liệu theo giá trị max, theo phân nguyên của giá trị chung bình, theo giá trị biên.

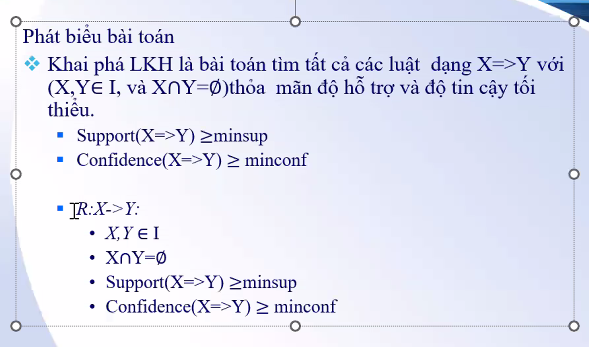




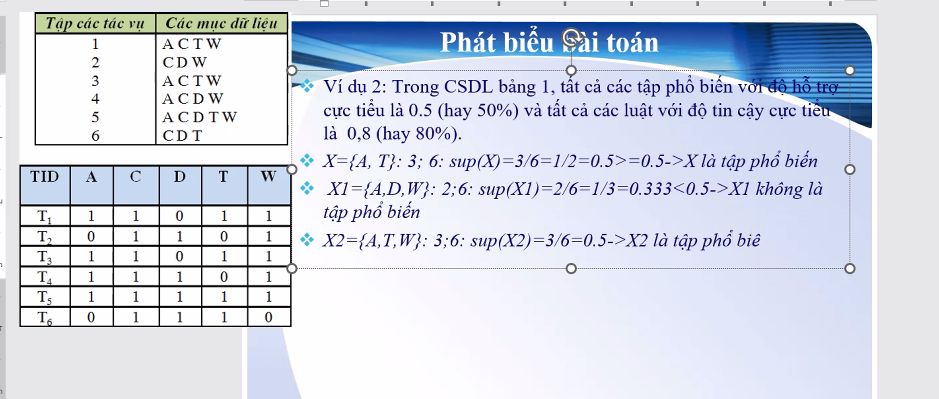
* Theo giá trị biên thì phải tính khoảng cách của giá trị đến biên, cái nào gần hơn thì lấy



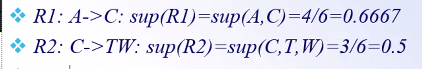
**Bài : Luật kết hợp**



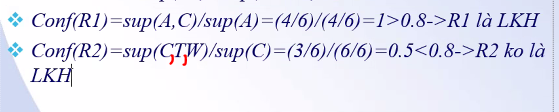
* Tìm luật phổ biến cho tập:



* Tìm tập phổ biến cho luật: (chỉ có tập phổ biến chứ k có luật phổ biến)



* Độ tin cậy của luật:



## I.Thuật giải Apriori

* Ý tưởng:
  + Tạo ra tập có 1 hạng mục: trong csdl có bao nhiêu hạng mục thì hình thành đc bấy nhiêu tập có 1 hạng mục, tính độ hỗ trợ của từng tập 1 hạng mục này, nếu > minsup thì lấy còn nhỏ hơn thì loại.

-> thu đc tập có 1 hạng mục là tập phổ biến.

* + Tạo tập có 2 hạng mục: tổ hợp từ tập phổ biến có 1 hạng mục -> tìm ra tập ứng viên -> tính độ hỗ trợ của từng tập ứng viên nếu > minsup thì lấy.

->Thu được tập có 2 hạng mục là tập phổ biến.

=> kết thúc khi tập ứng viên rỗng.

BT\_Luật kết hợp

## II. thuật giải AprioriTID

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| C1 | C1’ |  | C1 |  | L1 |  |
| item | TID | Các giao dịch | items | count | items | count |
| A | 100 | A,C,D,I | A | 4 | A | 4 |
| B | 200 | A,C,I | B | 3 | B | 3 |
| C | 300 | C,E,I | C | 3 | C | 3 |
| D | 400 | A,B,D,E | D | 4 | D | 4 |
| E | 500 | B,D,I | E | 3 | E | 3 |
| I | 600 | A,B,D,E | I | 4 | I | 4 |

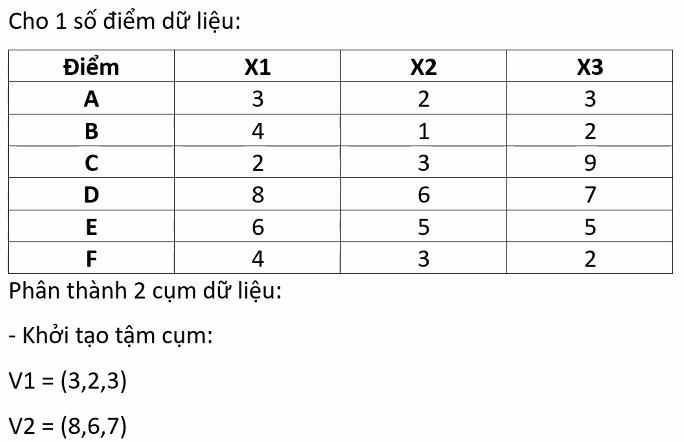
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| C2 | C2’ | | C2 | | L2 | |
| item | TID | Các giao dịch | items | count | items | count |
| AB | 100 | AC,AD,AI,CD,CI,DI | AB | 2 | AB | 2 |
| AC | 200 | AC,AI,CI | AC | 2 | AC | 2 |
| AD | 300 | CE,CI,EI | AD | 3 | AD | 3 |
| AE | 400 | AB,AD,AE,BD,BE,DE | AE | 2 | AE | 2 |
| AI | 500 | BD,BI,DI | AI | 2 | AI | 2 |
| BC | 600 | AB,AD,AE,BD,BE,DE | BC | 0 |  |  |
| BD |  |  | BD | 3 | BD | 3 |
| BE |  |  | BE | 2 | BE | 2 |
| BI |  |  | BI | 1 |  |  |
| CD |  |  | CD | 1 |  |  |
| CE |  |  | CE | 1 |  |  |
| CI |  |  | CI | 3 | CI | 3 |
| DE |  |  | DE | 2 | DE | 2 |
| DI |  |  | DI | 2 | DI | 2 |
| EI |  |  | EI | 1 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| C3 | C3’ | | C3 | | L3 | |
| item | TID | Các giao dịch  (khi tạo phải kiểm tra xem có trong C3 không nếu k có thì loại) | items | count | items | count |
| ABD | 100 | ACI,ADI | ABD | 2 | ABD | 2 |
| ABE | 200 | ACI | ABE | 2 | ABE | 2 |
| ACI | 300 |  | ACI | 2 | ACI | 2 |
| ADE | 400 | ABD, ABE,ABE,ADE,BDE | ADE | 2 | ADE | 2 |
| ADI | 500 |  | ADI | 1 |  |  |
| BDE | 600 | ABD, ABE,ABE,ADE,BDE | BDE | 2 | BDE | 2 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| C4 | C4’ | | C4 | | L4 | |
| item | TID | Các giao dịch  (khi tạo phải kiểm tra xem có trong C3 không nếu k có thì loại) | items | count | items | count |
| ABDE | 100 |  | ABDE | 2 | ABDE | 2 |
|  | 200 |  |  |  |  |  |
|  | 400 | ABDE |  |  |  |  |
|  | 600 | ABDE |  |  |  |  |

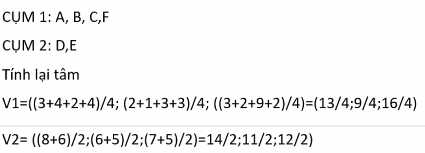
* Tập FI = {…}

Thuật toán k-mean

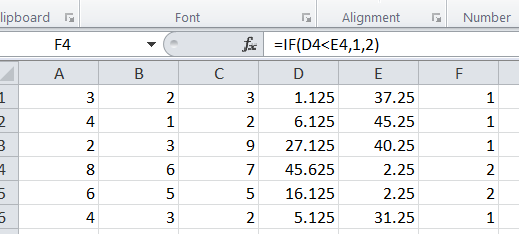


* Lặp lần 1:

|  |  |  |  |
| --- | --- | --- | --- |
|  | V1 | V2 |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| D |  | 0 |  |
| E |  |  |  |
| F |  |  | V1 |



* Lặp lần 2:

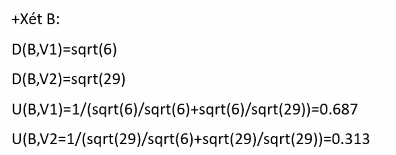


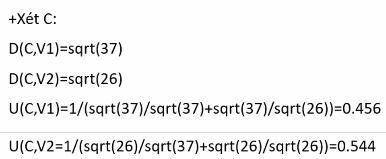
|  |  |  |  |
| --- | --- | --- | --- |
|  | V1 | V2 |  |
| A | = 1.06 |  | V1 |
| B |  |  | V1 |
| C |  |  | V1 |
| D |  |  | V2 |
| E |  |  | V2 |
| F |  |  | V1 |

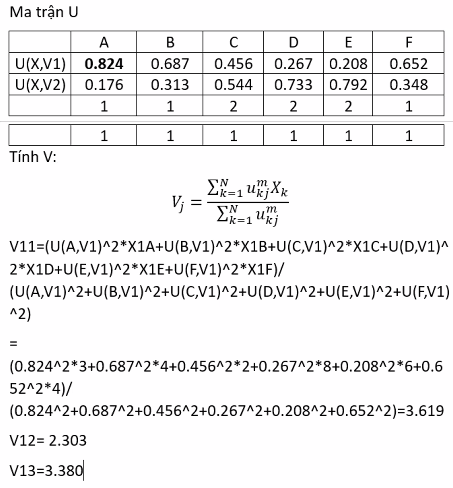
B2:

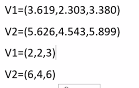


* Khởi tạo tâm ngẫu nhiên:

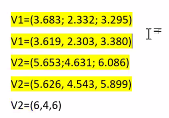








Lặp lần 2:

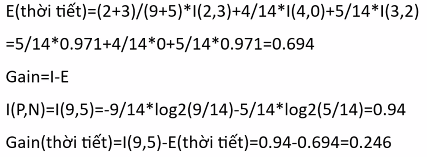


## ID3:



* Thời tiết

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Gía trị | P | N | I(P,N) |  |
| Nắng | 2 | 3 | 0.971 |  |
| U ám | 4 | 0 | 0 |  |
| mưa | 3 | 2 | 0.971 |  |



---

