**ΕΡΓΑΣΙΑ 4**

**ΣΤΟ ΜΑΘΗΜΑ: ΣΥΣΤΗΜΑΤΑ ΑΝΑΚΑΛΥΨΗΣ ΓΝΩΣΗΣ ΑΠΟ ΒΔ**

1. Αρχικά τρέχουμε τον αλγόριθμο που μας δίνεται με την στρατηγική Apriori και βρίσκουμε όλα τα συχνά στοιχειοσύνολα. Επίσης, επειδή έχουμε έξη συναλλαγές και το S = 0,33 τότε θα έχουμε Μinimum Support = 0,33\*6 = 1,98. Δηλαδή, θα κόβουμε τα στοιχειοσύνολα που έχουν count<2 (Αυτό φαίνεται και στους παρακάτω πίνακες).

**Στοιχεία (1-στοιχειοσύνολα)**

|  |  |
| --- | --- |
| **Item** | **Count** |
| Κραγιόν | 4 |
| Μάσκαρα | 2 |
| Ρούζ | 3 |
| Σκιές | 3 |
| Πούδρα | 4 |

**Στοιχεία (2-στοιχειοσύνολα)**

|  |  |
| --- | --- |
| **ItemSet** | **Count** |
| Κραγιόν, Μάσκαρα | 2 |
| Κραγιόν, Ρούζ | 2 |
| Κραγιόν, Σκιές | 2 |
| Κραγιόν, Πούδρα | 2 |
| ~~Μάσκαρα, Ρούζ~~ | ~~1~~ |
| ~~Μάσκαρα, Σκιές~~ | ~~0~~ |
| ~~Μάσκαρα, Πούδρα~~ | ~~0~~ |
| ~~Ρούζ, Σκιές~~ | ~~1~~ |
| Ρούζ, Πούδρα | 2 |
| Σκιές, Πούδρα | 3 |

**Στοιχεία (3-στοιχειοσύνολα)**

|  |  |
| --- | --- |
| **ItemSet** | **Count** |
| Κραγιόν, Σκιές, Πούδρα | 2 |
| ~~Κραγιόν, Ρούζ, Πούδρα~~ | ~~1~~ |

**ΣΗΜΕΙΩΣΗ**: Χρησιμοποιήσαμε την μέθοδο F(k-1) x F1.

Έπειτα, αφού καταλήξαμε ότι το πιο συχνό στοιχειοσύνολο είναι το {Κραγιόν, Σκιές, Πούδρα} μιας και εμφανίζεται δύο από τις έξη φορές, θα πάμε να δημιουργήσουμε τους κανόνες συσχέτισης (πρέπει να προκύψουν 2^κ-2 υποψήφιοι κανόνες συσχέτισης) δηλαδή τους παρακάτω έξη κανόνες:

* {Κραγιόν, Πούδρα} -> {Σκιές}

Confidence= {Κραγιόν, Σκιές, Πούδρα}/support{Κραγιόν, Πούδρα}\*100 = (2/2)\*100 = 100%

* {Κραγιόν, Σκιές} -> {Πούδρα}

Confidence= {Κραγιόν, Σκιές, Πούδρα}/support{Κραγιόν, Σκιές}\*100 = (2/2)\*100 = 100%

* {Σκιές, Πούδρα} -> {Κραγιόν}

Confidence= {Κραγιόν, Σκιές, Πούδρα}/support{Σκιές, Πούδρα}\*100 = (2/3)\*100 = 66%

* {Σκιές} -> {Κραγιόν, Πούδρα}

Confidence= {Κραγιόν, Σκιές, Πούδρα}/support{Σκιές}\*100 = (2/3)\*100 = 66%

* ~~{Πούδρα} -> {Κραγιόν, Σκιές}~~

Confidence= {Κραγιόν, Σκιές, Πούδρα}/support{Πούδρα}\*100 = (2/4)\*100 = 50%

* ~~{Κραγιόν} -> {Σκιές, Πούδρα}~~

Confidence= {Κραγιόν, Σκιές, Πούδρα}/support{Κραγιόν}\*100 = (2/4)\*100 = 50%

* Κραγιόν --> Μάσκαρα 2/4
* Μάσκαρα --> Κραγιόν 1
* Κραγιόν --> Πούδρα 2/4
* Πούδρα --> Κραγιόν 2/4
* Κραγιόν --> Ρουζ 2/4
* Ρουζ --> Κραγιόν 2/3
* Κραγιόν --> Σκιές 2/4
* Σκιές --> Κραγιόν 2/3
* Πούδρα --> Ρουζ 2/4
* Ρουζ --> Πούδρα 2/3
* Πούδρα --> Σκιές 3/4
* Σκιές --> Πούδρα 1

**ΣΗΜΕΙΩΣΗ**: Απορρίπτουμε τις τελευταίες δύο συσχετίσεις αφού έχουμε minconf=60% ενώ εμείς βγάλαμε και στις δυο περιπτώσεις confidence=50%<60%.Αρα θα δεχτούμε μόνο τα πρώτα τέσσερα.

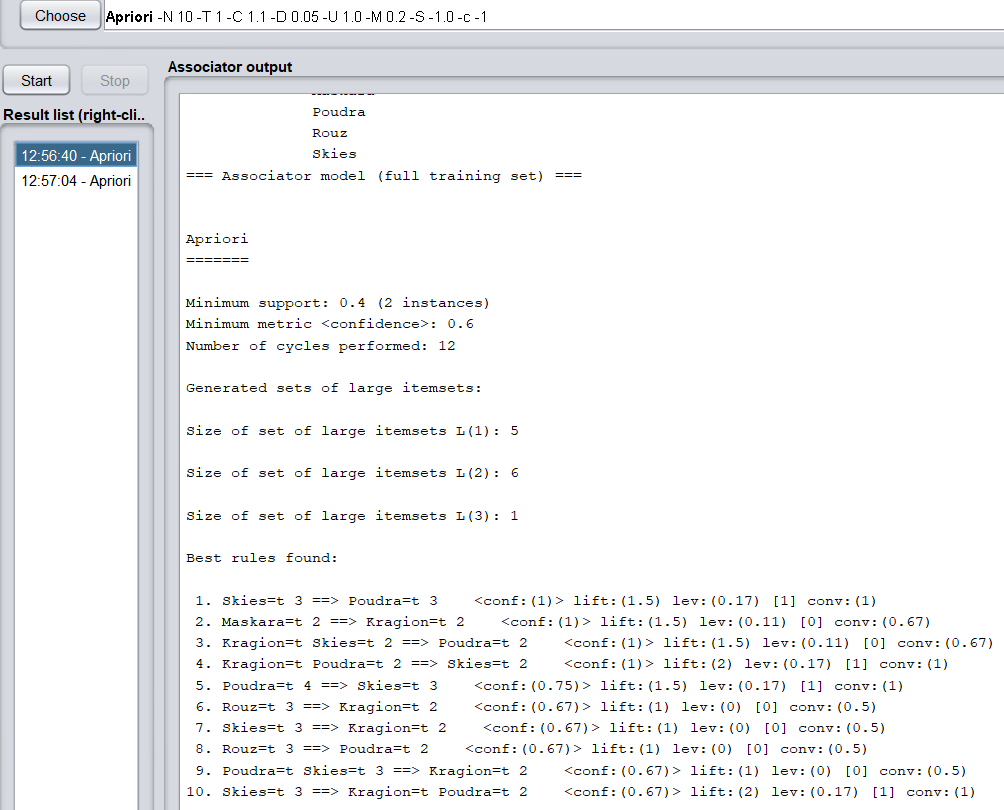
Τώρα πάμε να υπολογίσουμε τα lift (με τον τύπο **Lift(A,B) = Support(A,B) / (Support(A) x Support(B))** ) αυτών που δεχτήκαμε.

* **Lift(Κραγιόν, Πούδρα -> Σκιές) = 12/6 = 2 (Θετική συσχέτιση)**
* **Lift(Κραγιόν, Σκιές -> Πούδρα) = 12/8 = 1.5 (Θετική συσχέτιση)**
* **Lift(Σκιές, Πούδρα -> Κραγιόν) = 12/12 = 1 (Στατιστική ανεξαρτησία)**
* **Lift(Σκιές -> Κραγιόν, Πούδρα) = 12/6 = 2 (Θετική συσχέτιση)**
* Μάσκαρα --> Κραγιόν = 12/8 = 1.5
* Ρουζ --> Κραγιόν = 12/12 = 1
* Σκιές --> Κραγιόν = 12/12 = 1
* Ρουζ --> Πούδρα = 12/12 = 1
* Πούδρα --> Σκιές = 18/12 = 1.5
* Σκιές --> Πούδρα = 18/12 = 1.5

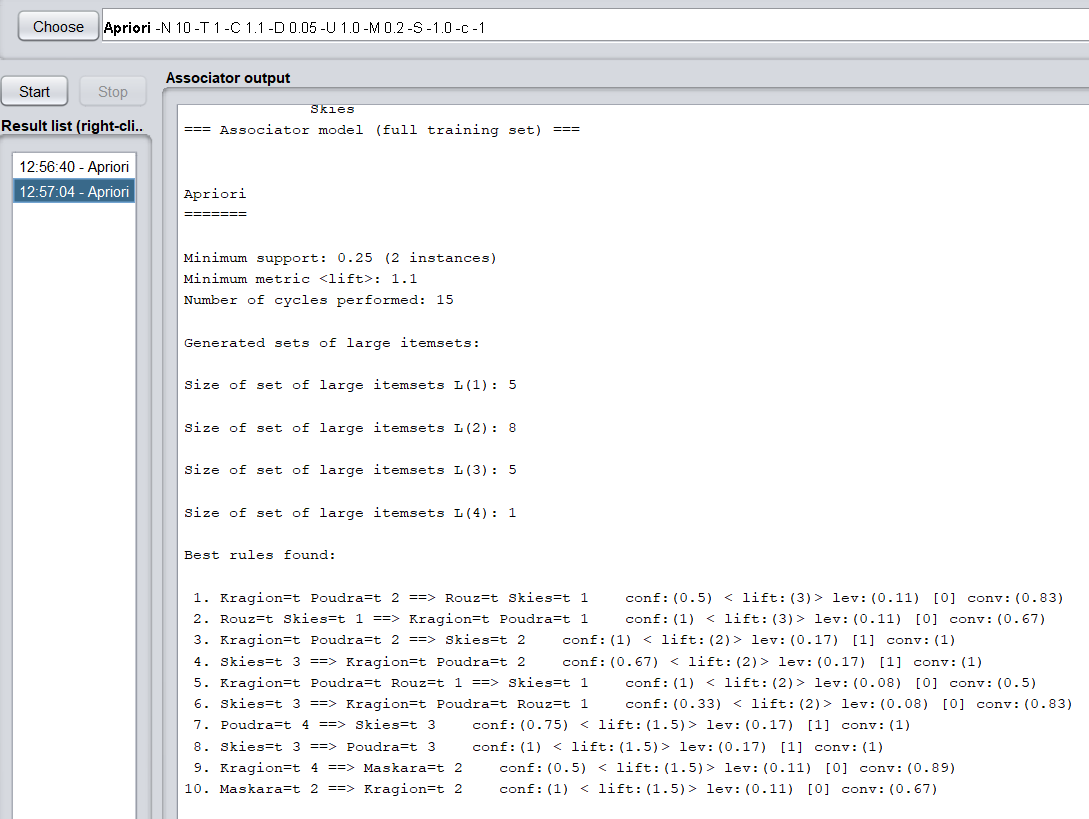
**ΣΧΟΛΙΑΣΜΟΣ:** Ο ποιο ισχυρός κανόνας είναι ο {Κραγιόν, Πούδρα -> Σκιές} αφού έχει c=100% και lift=2.

1. Πήγα και έτρεξα δυο φορές στο Weka Explorer τον Apriori και έθεσα και στις δυο περιπτώσεις lowerBoundMinSupport = 0,2. Στην πρώτη περίπτωση έθεσα ως **metric Type = Confidence** με **minMetric = 0,6** ενώ στην δεύτερη περίπτωση έθεσα ως **metric Type = Lift** και **minMetric = 0,6**. Εν τέλει πείρα τα παρακάτω αποτελέσματα:

**Περίπτωση 1 (confidence)**



**Περίπτωση 2 (lift)**



**ΣΧΟΛΙΑΣΜΟΣ:** Τα αποτελέσματα που πήραμε από το Weka όπως εύκολα διαπιστώνουμε συμβαδίζουν με τα αποτελέσματα του ερωτήματος 1.

**ΑΠΟ ΤΟΝ ΦΟΙΤΗΤΗ: Τσιαούση Σταύρο (dai17173)**