

Elyas Addawe – Saku Ihalainen – Hanna Kaimo

Fanituotteet

Tehtävä n:o 6

## 1 Tavoitteet

Kaupan johto haluaa saada selville minkäläisiä säännönmukaisuuksia on asiakkaiden ostokoreissa. Tavoitteena on saada selville, mitä tuotteita asiakkaat ostavat yhdessä. Kiinnostavinta on löytää tuoteryhmiä jotka ovat ylykkeenä toisen tuoteryhmän ostamiselle.

## 2 Data

Data sisältää tietoja asiakkaiden ostotapahtumista fanituotteita myyvässä verkkokaupassa. Data on tuotettu koneellisesti.

Attribuuttien määrä 20.

Attribuutit:

MovieBoxes: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

TShirts: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

CollegeShirts: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

Caps: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

Hats: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

Mugs: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

Music: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

Posters: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

Books: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

Scarves: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

PintAndWhiskeyGlasses: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

ScaleModels: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

CollectorCards: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

BoardGames: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

FlagsAndStandards: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

Stickers: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

Toys: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

Sweets: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

FridgeMagnets: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

UnderwearAndSocks: nominaali muuttuja ,arvot vain 0=ei ostettu tai 1=ostettu

Havaintojen määrä 100 000. Eli otanta on laaja.

Datassa ei ole puuttuvia tai poikkeavia arvoja.

### **3 Datan valmistelu**

Poistimme turhan ID- sarakkeen ja turhat välilyönnit datasta. Muokkasimme csv- muotoisen datan arff- muotoiseksi ja nimesimme attribuutit Wekaa varten myös niillä arvoilla, mitkä oli lueteltu alkuperäisessäkin datassa.

### **4 Mallinnus**

Datan assosiaatio sääntöjen etsimisessä käytimme Wekan Apriori- algoritmia. Se on ollut käytössä vuodesta 1994. Wekassa se löytyy kun klikkaa Associate palkkia. Muut Wekan assosiaatio- algoritmit ovat FPGrowth ja FilteredAssosiator.

Keskitymme positiivisen sääntöjen etsimiseen joten luovumme negatiivisen sääntöjen etsinnästä. Laitamme treatZeroAsMissing true valinnaksi. Tämä filttroi pois negatiiviset säännöt tulostuksista.

Alkutilanne tuki 0.1 eli 10 % ja luottamus 0.9,

Haluamme mahdollisimman korkean määrän tukea jotta olemme varmoja että kyseessä ei ole sattuma ja samalla parhaat säännönmukaisuudet saadaa saman tien.

```
=====
=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.1 (10000 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 10

Generated sets of large itemsets:

Size of set of large itemsets L(1): 14
Size of set of large itemsets L(2): 2

Best rules found:

1. FlagsAndStandards=1 11880 ==> Books=1 11145    <conf:(0.94)> lift:(4.73) lev:(0.09) [8786] conv:(12.94)
2. UnderwearAndSocks=1 14798 ==> FridgeMagnets=1 13476    <conf:(0.91)> lift:(4.42) lev:(0.1) [10423] conv:(8.88)
```

Löysimme vain 2 säännönmukaisuutta. Lippujen ja pöytäviirien ostaminen ennakoi kirjojen ostamista ja alusvaatteiden ja sukkien ostaminen jääkaappimagneettien ostamista.

Vähennämme luottamusta 0.6:een seuraavaksi. Muuten pidämme tilanteen ennallaan. Tulos on tämä:

```
Apriori
=====

Minimum support: 0.1 (10000 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 18

Generated sets of large itemsets:

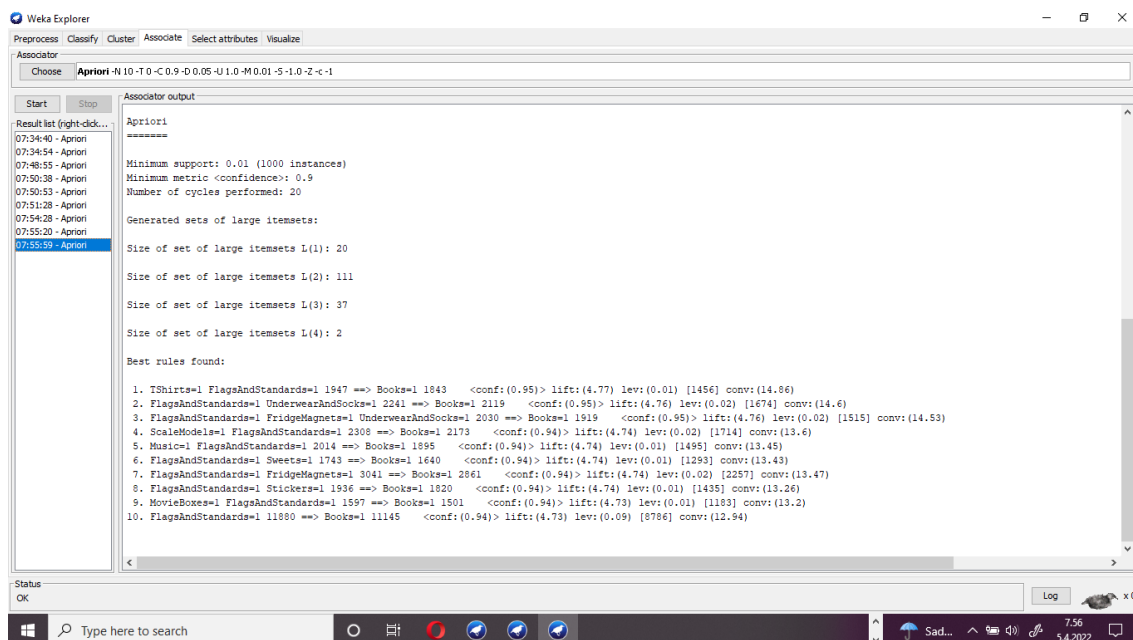
Size of set of large itemsets L(1): 14
Size of set of large itemsets L(2): 2

Best rules found:

1. FlagsAndStandards=1 11880 ==> Books=1 11145    <conf:(0.94)> lift:(4.73) lev:(0.09) [8786] conv:(12.94)
2. UnderwearAndSocks=1 14798 ==> FridgeMagnets=1 13476    <conf:(0.91)> lift:(4.42) lev:(0.1) [10423] conv:(8.88)
3. FridgeMagnets=1 20626 ==> UnderwearAndSocks=1 13476    <conf:(0.65)> lift:(4.42) lev:(0.1) [10423] conv:(2.46)
```

Vain säännön 2 käänteinen sääntö tulee lisää. Seuraavaksi vähennetään tuen määrää sillä ei saatu riittävän monta sääntöä datasta.

Tuki: 0.01 Luottamus: 0.9



Nyt assosiaatiosääntöjä saa valittua haluttu määrä, mutta tuki (0.01) on kymmenesosa default- tuen määrästä 0.1 eli 1 %.

Tuella 0.03 tulee myös kolme assosiaatiosääntöä.

Kun laitamme tuen arvon korkeaksi, esimerkiksi 0.4, Ei assosiaatiosääntöjä löydy.

```

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.15 -S -1.0 -Z -c -1
Relation: fanituotteet-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-weka.filters.unsupervised.attribute
Instances: 100000
Attributes: 20
MovieBoxes
TShirts
CollegeShirts
Caps
Hats
Mugs
Music
Posters
Books
Scarves
PintsAndWhiskeyGlasses
ScaleModels
CollectorCards
BoardGames
FlagsAndStandards
Stickers
Toys
Sweets
FridgeMagnets
UnderwearAndSocks
=== Associator model (full training set) ===

No large itemsets and rules found!

```

Luottamuksen vähentäminen ei lisää tuloksia kovin paljon, esimerkiksi kun luottamus on 0.2, tulee 4 assosiaatiosääntöä.

Weka tulkitsee kysymysmerkit puuttuviksi arvoiksi, joten niiden käyttäminen Wekassa nollien tilalla datassa johtaa samaan lopputulokseen kuin treatZeroAsMissing-parametrin asettaminen arvoon true.

Jos vaihdamme treatZeroAsMissing- parametrin arvoksi false, assosiaatiosääntöjä kyllä löytyy, mutta ne ilmentävät sitä, että jos et ostanut tätä, et ostanut tuotakaan ja sitä ei tässä toimeksiannossa pyydetty selvittämään.

```
Minimum support: 0.9 (90000 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 1

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Size of set of large itemsets L(2): 10

Size of set of large itemsets L(3): 6

Best rules found:

1. CollegeShirts=0 Caps=0 93281 ==> CollectorCards=0 92073    <conf:(0.99)> lift:(1) lev:(0) [41] conv:(1.03)
2. Caps=0 Mugs=0 93596 ==> CollectorCards=0 92377    <conf:(0.99)> lift:(1) lev:(0) [34] conv:(1.03)
3. Caps=0 96415 ==> CollectorCards=0 95154    <conf:(0.99)> lift:(1) lev:(0) [29] conv:(1.02)
4. CollegeShirts=0 Mugs=0 93884 ==> CollectorCards=0 92645    <conf:(0.99)> lift:(1) lev:(0) [18] conv:(1.01)
5. CollegeShirts=0 Toys=0 91309 ==> CollectorCards=0 90098    <conf:(0.99)> lift:(1) lev:(0) [11] conv:(1.01)
6. CollegeShirts=0 96729 ==> CollectorCards=0 95446    <conf:(0.99)> lift:(1) lev:(0) [12] conv:(1.01)
7. Mugs=0 Toys=0 91633 ==> CollectorCards=0 90413    <conf:(0.99)> lift:(1) lev:(0) [6] conv:(1)
8. Mugs=0 97048 ==> CollectorCards=0 95755    <conf:(0.99)> lift:(1) lev:(0) [6] conv:(1)
9. Toys=0 94382 ==> CollectorCards=0 93117    <conf:(0.99)> lift:(1) lev:(-0) [-1] conv:(1)
10. CollectorCards=0 Toys=0 93117 ==> Mugs=0 90413    <conf:(0.97)> lift:(1) lev:(0) [44] conv:(1.02)
11. Toys=0 94382 ==> Mugs=0 91633    <conf:(0.97)> lift:(1) lev:(0) [37] conv:(1.01)
12. CollegeShirts=0 Caps=0 93281 ==> Mugs=0 90563    <conf:(0.97)> lift:(1) lev:(0) [35] conv:(1.01)
13. Caps=0 CollectorCards=0 95154 ==> Mugs=0 92377    <conf:(0.97)> lift:(1) lev:(0) [31] conv:(1.01)
14. Caps=0 96415 ==> Mugs=0 93596    <conf:(0.97)> lift:(1) lev:(0) [27] conv:(1.01)
15. CollegeShirts=0 CollectorCards=0 95446 ==> Mugs=0 92645    <conf:(0.97)> lift:(1) lev:(0) [16] conv:(1.01)
16. CollegeShirts=0 96729 ==> Mugs=0 93884    <conf:(0.97)> lift:(1) lev:(0) [10] conv:(1)
```

## 5 Arviointi

Korkealla tuella assosiaatiosääntöjä tulee vain kaksi ja matalalla tuella assosiaatiot voivat johtua sattumasta eli niihin ei voi luottaa. Nämä asiat selittyvät sillä, minkä datastakin huomaa: Fanituotteiden myynti on heikkoa. Arvoja 0, asiakas ei ole ostanut tuotetta on jokaisessa attribuutissa todella paljon.

Tämä selittää myös sen, että kun treatZeroAsMissing- arvo on true, tulee ensimmäisiksi assosiaatiosäännöiksi 0 arvojen assosiaatioita.

Aineistoista esiintyi kaksi hyvää säännönmukaisuutta jossa on yli 0.9 luottamus ja 0.1 tuki. Flags And Standards johtaa books tuoteryhmän ostamiseen. Underwear and Socks johtaa fridgemagnet tuoteryhmän ostamiseen. Mutta tuki arvolla 0.1 ei ole sekään erityisen korkea.

Löydetyt assosiaatiosäännöt ovat yllättäviä, esimerkiksi alusvaatteiden ja sukkien ostajat ostavat jääkaappimagneetteja. Suoraa loogista selitystä assosiaatioihin on vaikea sanoa.

## **6 Täytäntöönpano**

Vaikea sanoa myös tämän tutkimuksen perusteella tuoteyhdistelmiä, joita kannattaisi tehdä, jotka menisivät kaupaksi. Toimenpidesuosituksena esimerkiksi vähentää tuotevalikoimaa, keskittyä muutamaan tuotteeseen, joiden joukossa olisivat ainakin ne, joita menee eniten kaupaksi. Nostaa brändin suosiota ihmisten keskuudessa, jolloin verkkokaupan fanituotteitakin menisi ehkä enemmän kaupaksi.