

Timo Harju – Samuli Schroderus – Pasi Saarela – Joonas
Lehikoinen

Analyysiraportti (CRISP-DM) Tehtävä no: 6

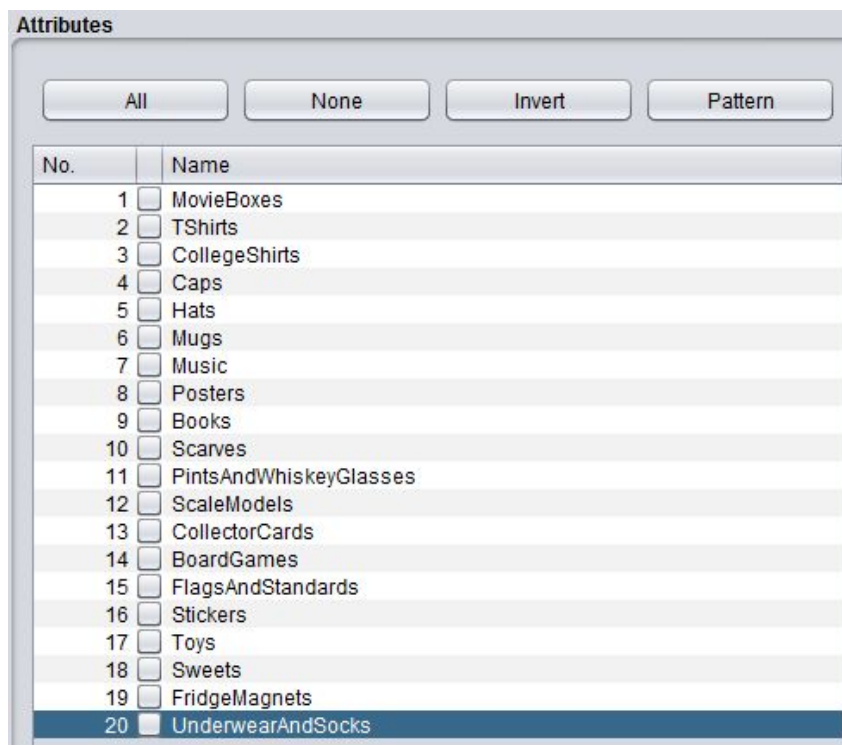
Tehtävä n:o 1

1 Tavoitteet

Tavoitteena on selvittää myyntitietokannan sisältävien ostotapahtumien perusteella, säännönmukaisuuksia asiakkaan ostoskäyttäytymisessä. Aineisto koostuu sadastatuhannesta ostotapahtumasta yhtenäiseltä ajanjaksolta. Tutkitaan kuinka eri Apriori algoritmin parametrien muuttaminen vaikuttaa saatuihin tuloksiin ja lopuksi todetaan tulokset.

2 Data

Data sisältää 100 000 ostotapahtumaa, jotka sisältävät tuoteryhmiä kuvaavia attribuutteja numeraalisessa muodossa, joko nolla (0) tai yksi (1) arvoja jokaisen kohdalla liittyen siihen onko tuotetta ostettu vai onko tuote jäänyt ostamatta ostotapahtumassa.



The screenshot shows a window titled "Attributes" with four buttons at the top: "All", "None", "Invert", and "Pattern". Below these buttons is a table with two columns: "No." and "Name". The table lists 20 product categories, each with a checkbox in the "No." column. The categories are: MovieBoxes, TShirts, CollegeShirts, Caps, Hats, Mugs, Music, Posters, Books, Scarves, PintsAndWhiskeyGlasses, ScaleModels, CollectorCards, BoardGames, FlagsAndStandards, Stickers, Toys, Sweets, FridgeMagnets, and UnderwearAndSocks. The "UnderwearAndSocks" row is highlighted in blue.

No.	Name
1	<input type="checkbox"/> MovieBoxes
2	<input type="checkbox"/> TShirts
3	<input type="checkbox"/> CollegeShirts
4	<input type="checkbox"/> Caps
5	<input type="checkbox"/> Hats
6	<input type="checkbox"/> Mugs
7	<input type="checkbox"/> Music
8	<input type="checkbox"/> Posters
9	<input type="checkbox"/> Books
10	<input type="checkbox"/> Scarves
11	<input type="checkbox"/> PintsAndWhiskeyGlasses
12	<input type="checkbox"/> ScaleModels
13	<input type="checkbox"/> CollectorCards
14	<input type="checkbox"/> BoardGames
15	<input type="checkbox"/> FlagsAndStandards
16	<input type="checkbox"/> Stickers
17	<input type="checkbox"/> Toys
18	<input type="checkbox"/> Sweets
19	<input type="checkbox"/> FridgeMagnets
20	<input type="checkbox"/> UnderwearAndSocks

Kuva 1: attribuutit

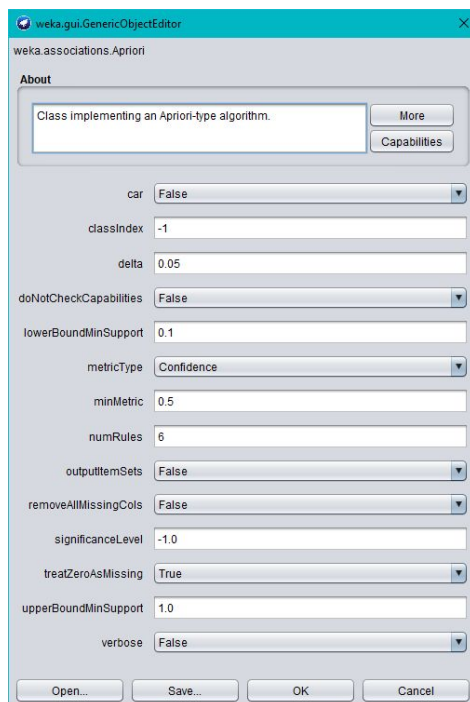
3 Datan valmistelu

Aineistosta poistettiin ID-attribuutti ja muut attribuutit muutettiin numeraalisista muuttujista luokkamuuttujiksi Wekan NumericToNominal filttterillä.

4 Mallinnus

Assosiaatioalgoritmina käytettiin Aprioria. Algoritmin parametreinä kokeiltiin eri arvoja, joiden arvot ovat nähtävissä alla olevissa kuvissa (Malli 1. ja Malli 2.).

Malli 1:



Ensimmäiseen malliin valittiin ostotapahtumien minimimääräksi 10 000 tapausta asettamalla lowerBoundMinSupport parametri arvoon 0.1. Confidence määritettiin arvoon 0,5 ja etsittävien sääntöjen määräksi 6 asettamalla numRules parametri arvoon 6.

Parametrit valittiin, jotta merkittävimmät säännöt saataisiin selville.

Näillä parametreillä saatiin neljä erilaista sääntöä.

```

Apriori
=====

Minimum support: 0.1 (10000 instances)
Minimum metric <confidence>: 0.5
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)
4. Books=1 19853 ==> FlagsAndStandards=1 11145 <conf:(0.56)> lift:(4.73) lev:(0.09) [8786] conv:(2.01)

```

Malli 1. minsupport 0.1 (10 000 tuotetta vähintään), minmetric 0.5 (vähintään 0.5 confidence), numrules 6 (enintään 6 parasta sääntöä), upperboundMinSupport 1.0 (maksimissaan 100 000 tuotetta)

Malli 2:

Toiseen malliin valittiin ostotapahtumien minimimääräksi 2000 tapausta asettamalla lowerBoundMinSupport parametri arvoon 0.02. Confidence määritettiin arvoon 0,93 ja etsittävien sääntöjen määräksi 6 asettamalla numRules-parametri arvoon 6.

Kyseiset parametrit valittiin, jotta sääntöjä saataisiin esille useampia.

Näillä parametreilla saatiin viisi erilaista sääntöä.

```
Apriori
=====

Minimum support: 0.02 (2000 instances)
Minimum metric <confidence>: 0.93
Number of cycles performed: 20

Generated sets of large itemsets:

Size of set of large itemsets L(1): 19

Size of set of large itemsets L(2): 65

Size of set of large itemsets L(3): 13

Best rules found:

1. FlagsAndStandards=1 UnderwearAndSocks=1 2241 ==> Books=1 2119    <conf:(0.95)> lift:(4.76) lev:(0.02) [1674] conv:(14.6)
2. ScaleModels=1 FlagsAndStandards=1 2308 ==> Books=1 2173    <conf:(0.94)> lift:(4.74) lev:(0.02) [1714] conv:(13.6)
3. FlagsAndStandards=1 FridgeMagnets=1 3041 ==> Books=1 2861    <conf:(0.94)> lift:(4.74) lev:(0.02) [2257] conv:(13.47)
4. FlagsAndStandards=1 11880 ==> Books=1 11145    <conf:(0.94)> lift:(4.73) lev:(0.09) [8786] conv:(12.94)
5. Posters=1 FlagsAndStandards=1 2352 ==> Books=1 2195    <conf:(0.93)> lift:(4.7) lev:(0.02) [1728] conv:(11.93)
```

Malli 2. minsupport 0.02 (2000 tuotetta vähintään), minmetric 0.93 (vähintään 0.93 confidence), numrules 6 (enintään 6 parasta sääntöä), upperboundMinSupport 1.0 (maksimissaan 100 000 tuotetta)

5 Arviointi

Mallin 1. tuloksia tutkittaessa huomattiin sääntöjen 3. ja 4. luottamusarvon olevan alhaiseita (0,65 ja 0,56) verrattaessa sääntöihin 1. ja 2. (0.94, 0.91). Tästä voitiin päätellä, että ostopäätöksen todennäköisyyden kannalta kannattaisi huomioida vain säännöt 1. ja 2. Vähäinen sääntöjen määrä ja suuri luottamus arvojen välinen ero toisistaan selittyy ostotapahtumien minimi määrän asettamisesta 10000 (LowerboundMinSupport=0,1).

Mallin 2. tuloksista voidaan todeta, että huomattavan usein yhdessä myyviä tuotepareja löytyy useita, vaikka niiden myyntivolumi ei ole kaupankäynnin kannalta kovin merkittävä.

6 Täytäntöönpano

Myyntin/tuotesijoittelun kannalta isommalla volyymilla tapahtuvat ostotapahtumien kausaliitteit ovat oleellisempia, vaikka ne eivät olisi yhtä todennäköisiä kuin esimerkiksi pienemmällä volyymilla tapahtuvat ostot. Tavoitteena on myyntin kasvattaminen, jolloin kuvan 3 sääntöihin ja syy-seuraus suhteisiin kannattaisi fokusoida enemmän, kuin kuvan 2 tapahtuviin vähäisemmän volyymin assosiaatioihin.

Analyysissä selvinneiden sääntöjen perusteella suosittelemme myynnin edistämiseksi seuraavaa: kirjat, julisteet ja pienoismallit kannattaa sijoittaa lippujen läheisyyteen, sekä magneetit alusvaatteiden ja sukkien yhteyteen myymälässä ja mainonnassa.