

Asociácia produktov a profilácia zákazníckych skupín pomocou dolovania dát

Products Associations and Customer Groups Profiling Using Data Mining

Andrej Trnka

Abstract: The paper points out to the use of association rules in marketing through data mining. Marketing sector provides large potential of applying various data mining methods and algorithms. One of them is the use of association rules in the so-called shopping cart analysis. These associations are able to identify relationships between the products that the customer buys. The results then allow determining the sequences, which can help in better selling of products. In case of the availability of data concerning customers it is also possible to perform profiling the customer groups.

Key words: association rules, market basket analysis, data mining, marketing.

Abstrakt: Článok poukazuje na využitie asociačných pravidiel v marketingu pomocou dolovania dát. Marketingový sektor poskytuje veľké možnosti nasadenia rozličných metód a algoritmov dolovania dát. Jednou z nich je použitie asociačných pravidiel v tzv. analýze nákupného košíka. Tieto asociácie dokážu identifikovať vzťahy medzi produktmi, ktoré spotrebiteľ nakupuje. Z výsledkov je potom možné stanoviť postupnosti, ktoré môžu dopomôcť k lepšiemu predaju produktov. V prípade dostupnosti dát týkajúcich sa spotrebiteľov je možné vykonať tiež profiláciu skupín spotrebiteľov.

Kľúčové slová: asociačné pravidlá, analýza nákupného košíka, dolovanie dát, marketing.

1. Analýza nákupného košíka

Cieľom analýzy nákupného košíka je identifikovať produkty alebo skupiny produktov, ktoré majú tendenciu vyskytovať sa spoločne (sú asociované) v nákupnej transakcii. Inak povedané, v nákupnom košíku. Poznatky získané z analýzy nákupného košíka môžu byť veľmi cenné. Napríklad zamestnanci supermarketov môžu reorganizovať rozloženie tovaru tak, aby

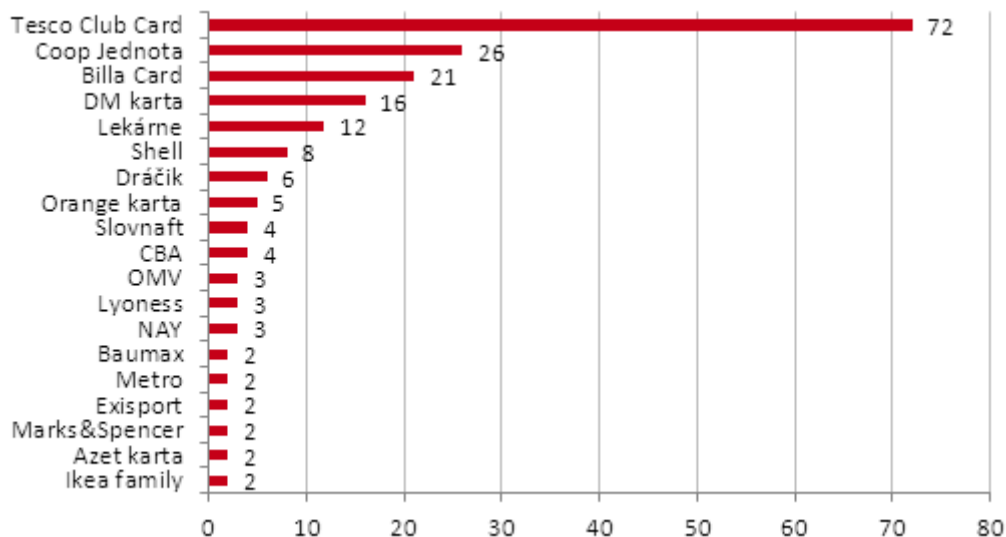
produkty ktoré sa predávajú spolu boli umiestnené blízko seba. Iné využitie je na zlepšenie účinnosti propagačnej kampane – produkty ktoré sú asociované nesmú byť súčasne v kampani. Stačí, keď bude v kampani podporovaný len jeden výrobok a následne môžeme predpokladať, že sa zvýši aj predaj asociovaných výrobkov.

Dátový set ktorý sa zvyčajne používa pri analýze nákupného košíka obsahuje dáta z transakcií vykonaných v určitom období (napr. jeden rok) a v určitých predajných miestach (napr. reťazce supermarketov). Keďže spotrebiteľ nakupuje nie len jeden krát v sledovanom období, v dátovom sete sa vyskytuje viackrát. V skutočnosti sa spotrebiteľ vyskytne v databáze vždy, keď vykoná transakciu na predajnom mieste. Z týchto dát je možné zistiť najčastejšiu kombináciu produktov zakúpených spotrebiteľom. Asociačné pravidlá predstavujú najprirodzenejšiu metodiku na zistenie týchto kombinácií a boli vyvinuté práve pre tento účel. Analýza kombinácií produktov, ktoré spotrebiteľ nakúpil a počet opakovaní týchto kombinácií vedie k vytvoreniu asociačných pravidiel. Každé takéto pravidlo popisuje konkrétnu vzorku. Súbor asociačných pravidiel môže byť veľmi jednoducho interpretovaný. Najčastejším spôsobom použitým pri zbere dát sú vernostné karty, ktoré sú veľmi rozšírené aj v našich končinách. (1)

2. Vernostné karty

Pomocou vernostnej karty môže byť vykonaná analýza, za ktorú by musel obchodník za iných okolností zaplatiť. Aj preto sa prostredníctvom zliav rád vzdá časti svojho zisku. Kartičku najväčšieho obchodného reťazca na Slovensku, Tesco Stores si napríklad za rok 2010 zaobstaralo 870 000 ľudí. (2)

Vernostné karty niektorej zo spoločností v súčasnosti vlastní až 83 % slovenskej populácie. Vyplýva to z prieskumu agentúry TNS Slovakia, ktorý sa uskutočnil v júli tohto roku na reprezentatívnej vzorke 500 respondentov. Vernostné karty pritom vlastní predovšetkým ženy a obyvatelia Bratislavského kraja. Medzi najvyužívanejšie vernostné karty patria karty obchodných reťazcov, v ktorých ľudia najčastejšie nakupujú. Približne tri štvrtiny všetkých vlastníkov vernostných kariet využíva kartu Club Card obchodného reťazca Tesco. S výrazným rozdielom nasleduje využívanie ostatných lojalitných kariet. Štvrtina opýtaných používa kartu Coop Jednota a pätina Billa kartu. Nasledovala DM karta s podielom 16 %, vernostné karty lekární s podielom 12 %, karta Shell s 8 %, Dráčik a Dr. Max po 6 % a Orange s 5 % (obr. 1). V priemere vlastní Slováci tri vernostné karty. (3), (4)



Obrázok 1: Podiel vernostných kariet na Slovensku

(Zdroj: TNS Slovakia)

3. Spracovanie dát z vernostných kariet

Pri spracovaní a vyhodnocovaní dát môžeme použiť dve metódy dolovania dát:

- asociačné pravidlá (algoritmus Apriori) spolu s pavučinovým diagramom odhaľujúcim väzby medzi zakúpenými produktmi,
- algoritmus C5.0 na vytvorenie profilácie spotrebiteľov identifikovaných skupín produktov.

V dátovom sete, ktorý sme použili boli zozbierané nasledovné údaje o spotrebiteľoch a nakupovaných produktoch:

- sumárne informácie o nákupe,
- informácie o spotrebiteľovi,
- obsah nákupného košíka.

Použitý dátový set obsahuje 634 záznamov a pre jednoduchosť použijeme iba 10 produktov. V skutočnosti ich však môže byť oveľa viac. Pre väčšie objemy dát je vhodné vytvorenie dátového skladu, do ktorého sa potom v určitých časových intervaloch zavádzajú zozbierané dáta. (5), (6)

Ako prvý krok analýzy použijeme tzv. Apriori algoritmus. Tento algoritmus generuje asociačné pravidlá, pomocou ktorých môžeme analyzovať vzťahy (asociácie) medzi produktmi v nákupnom košíku. Do tejto analýzy vstupujú len tie dáta z dátového setu, ktoré sa týkajú iba produktov. Čiže sumárne informácie o nákupe a informácie o spotrebiteľovi môžeme vyfiltrovať a tým zrýchliť výpočtovú náročnosť algoritmu.

Výsledok Apriori algoritmu obsahuje asociačné pravidlá, zobrazené v tabuľke 1.

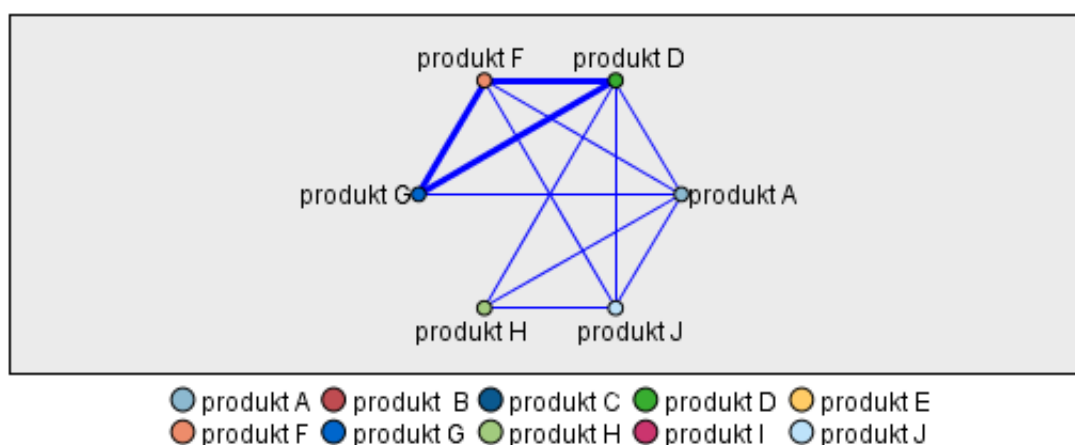
Tabuľka 1: Vytvorené asociačné pravidlá

Consequent	Antecedent	Support %	Confidence %
produkt F	produkt G produkt D	16,562	88,571
produkt D	produkt G produkt F	17,035	86,111
produkt G	produkt F produkt D	17,035	86,111

(Zdroj: vlastné spracovanie)

V minulosti bolo možné použiť na vytvorenie asociačných pravidiel algoritmus GRI, ktorý bol nahradený použitým algoritmom Apriori. Ten vytvorí asociačné pravidlá iba medzi silne asociovanými produktmi, t.j. zvyšuje prehľadnosť ponúknutého výstupu.

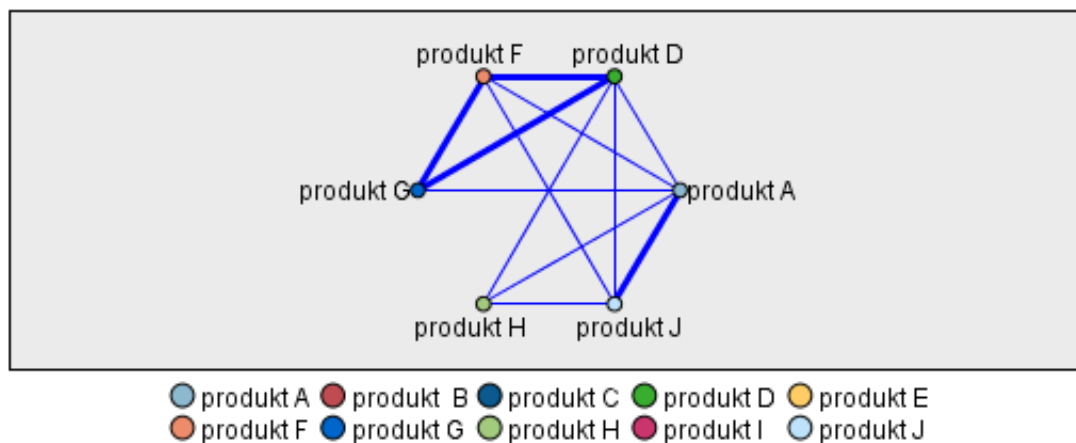
Z tabuľky 1 je vidieť, že asociačné pravidlá zobrazujú rôzne asociácie medzi produktom F, produktom D a produktom G. Prítomnosť obojsmerných asociačných pravidiel (napr. produkt F – produkt D a produkt D a produkt F) naznačuje možnosť použitia tzv. pavučinového grafu. Použitie tohto typu grafu môže pomôcť lepšie pochopiť asociácie medzi produktmi (obr. 2).



Obrázok 2: Pavučinový graf asociácií produktov

(Zdroj: vlastné spracovanie)

Pretože sa v nákupných košíkoch nachádza väčšina kombinácií produktov, vhodným nastavením parametrov grafu je možné pozorovať viaceré asociácie zobrazené hrubou čiarou (obr. 3). Tieto asociácie dopomáhajú k tomu, aby sme pomocou nich mohli identifikovať určité skupiny spotrebiteľov navrhnuté vytvoreným modelom.



Obrázok 3: Upravený pavučinový graf asociácií produktov

(Zdroj: vlastné spracovanie)

Z obr. 3 je vidieť, že vynikajú dve skupiny spotrebiteľov:

- ktorí kupujú produkt D, produkt F a produkt G (skupina A),
- ktorí kupujú produkt A a produkt J (skupina B).

Poslednou časťou analýzy je profilácia spotrebiteľov, ktorí patria do týchto výrazných skupín.

4. Profilácia spotrebiteľských skupín

Po identifikovaní hlavných skupín spotrebiteľov podľa produktov ktoré nakupujú, je dobré vedieť o akých spotrebiteľov ide. Čiže chceme spoznať ich demografický profil. Na realizáciu tohto kroku sme použili algoritmus C5.0, ktorý generuje pravidlá na začlenenie do skupín. Pred samotným použitím algoritmu sme museli odvodiť nové premenné pre každú skupinu spotrebiteľov. Tieto premenné boli následne pridané do dátového setu a pri každom zázname určovali, či spotrebiteľ patrí do jednej alebo druhej skupiny.

Pred samotnou profiláciou spotrebiteľských skupín bolo potrebné zmeniť dáta vstupujúce do analýzy. Tentoraz sme použili demografické dáta, ktoré boli určené ako vstupné a za cieľové dáta boli určené odvodené premenné z predchádzajúceho kroku.

Ako výsledok profilácie spotrebiteľských skupín vystupuje zoznam if-then pravidiel, ktoré určujú začlenenie do jednotlivých skupín.

Pre skupinu A môžeme konštatovať, že najdôležitejšie prediktory sú premenné príjem a pohlavie.

Pravidlo pre hodnotu T (patrí do skupiny A):

```
Rule 1 for T
  if pohlavie = M
  and príjem <= 16 900
  then T
```

Pravidlá pre hodnotu F (nepatrí do skupiny A):

```
Rule 1 for F
  if príjem > 16 900
  then F
```

```
Rule 2 for F
  if pohlavie = F
  then F
```

Pre skupinu B môžeme konštatovať, že najdôležitejšie prediktory sú premenné vek, príjem, cena nákupu a metóda platby.

Pravidlá pre hodnotu T (patrí do skupiny B):

```
Rule 1 for T
  if cena nákupu > 46,367
  and príjem <= 21 100
  and vek > 20
  and vek <= 24
  then T
```

```
Rule 2 for T
  if metóda platby = CASH
  and príjem <= 21 100
  and vek > 20
  and vek <= 24
  then T
```

```
Rule 3 for T
```

if príjem > 26 800

and vek <= 24

then T

Rule 4 for T

if vek <= 16

then T

Rule 5 for T

if príjem <= 21 100

and vek > 20

and vek <= 24

then T

Pravidlá pre hodnotu F (nepatrí do skupiny B):

Rule 1 for F

if vek > 24

then F

Rule 2 for F

if cena nákupu <= 46,367

and metóda platby = CARD

and vek > 20

then F

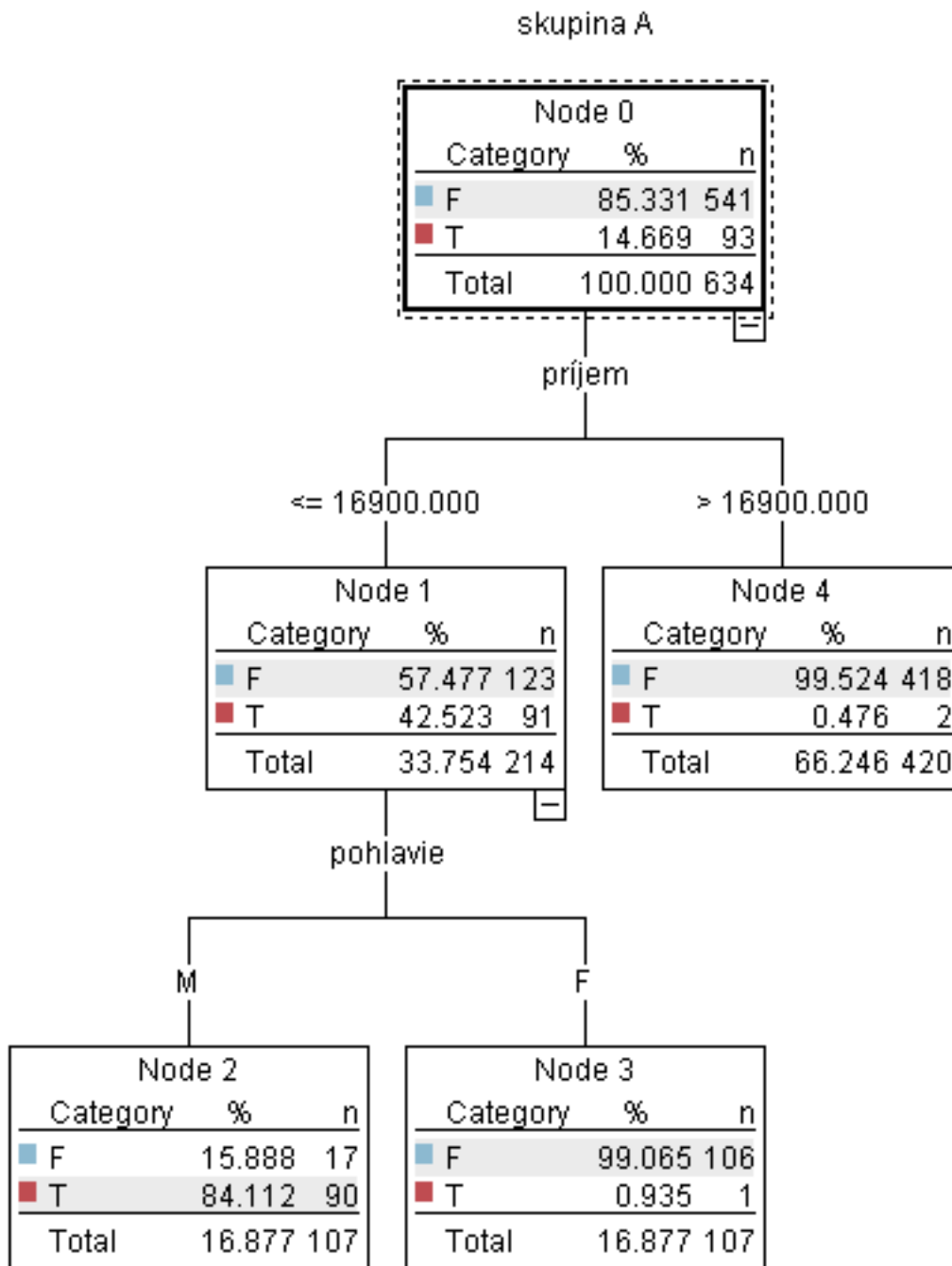
Rule 3 for F

if príjem <= 26 800

and vek > 16

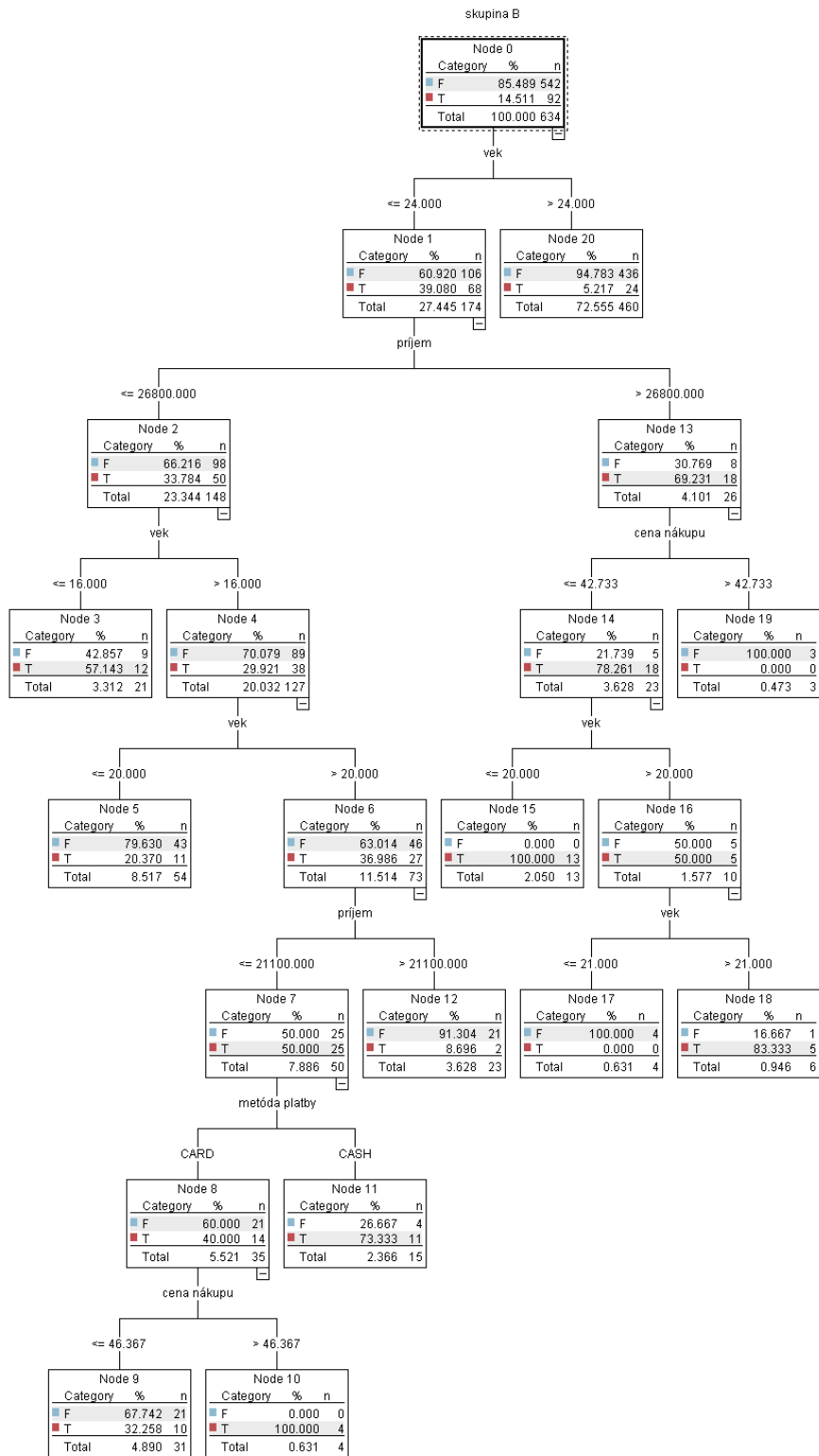
then F

Pre vizualizáciu predchádzajúcich pravidiel môžeme ešte použiť zobrazenie pomocou rozhodovacích stromov. Na obr. 4 je zobrazený rozhodovací strom pre zaradenie do skupiny A a na obr. 5 pre zaradenie do skupiny B.



Obrázok 4: Rozhodovací strom pre profilovanie spotrebiteľov pre skupinu A

(Zdroj: vlastné spracovanie)



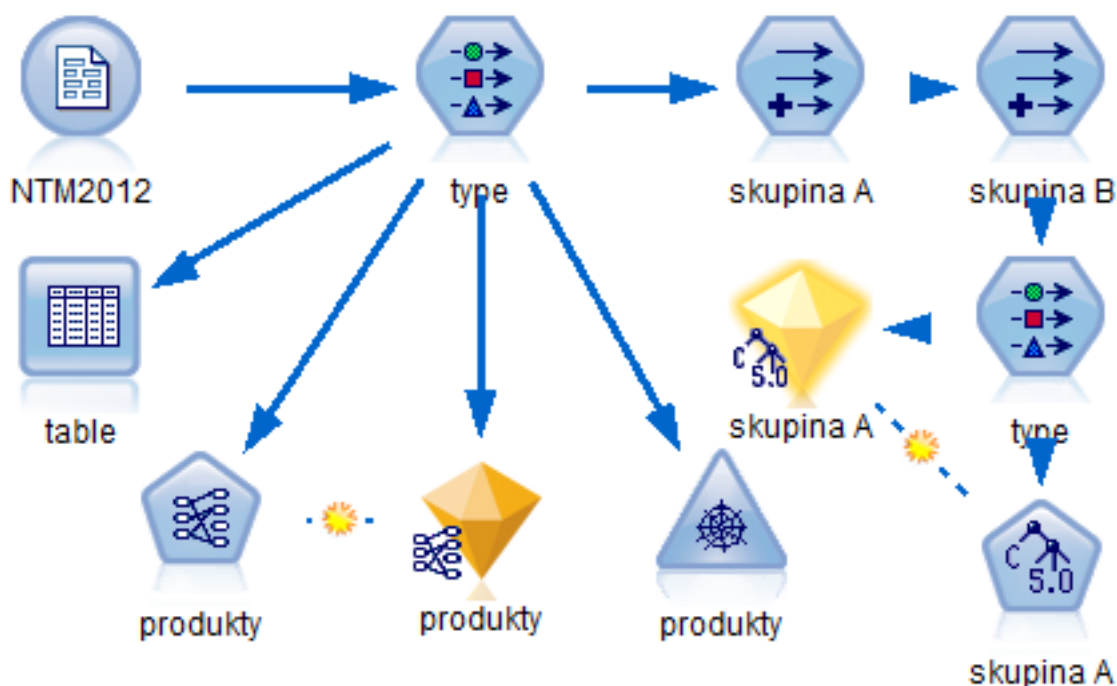
Obrázok 5: Rozhodovací strom pre profilovanie spotrebiteľov pre skupinu B

(Zdroj: vlastné spracovanie)

5. Záver

Článok uvádza možnosti aplikovania metód dolovania dát do marketingu. V prvej časti popisuje možnosti odhalenia asociácií medzi jednotlivými produktmi, ktoré spotrebiteľia nakupujú. Na zistenie asociačných pravidiel bol použitý algoritmus Apriori a následne pavučinový graf. Z výsledkov analýzy bolo možné špecifikovať dve veľké skupiny spotrebiteľov. V druhej časti článku boli tieto skupiny profilované. K tomu poslúžili demografické údaje o spotrebiteľoch. Výsledok profilácie bol zobrazený pomocou pravidiel if-then alebo pomocou rozhodovacieho stromu. Takto získané informácie je možné použiť na lepšie zacielenie marketingovej kampane. Veľmi diskutovanou témou je ochrana osobných údajov poskytovaných spotrebiteľom, čo je však nad rámec tohto článku.

Model, ktorý sme vytvorili a v ktorom sme vykonali analýzy je zobrazený na obr. 6.



Obrázok 6: Vytvorený model dolovania dát

(Zdroj: vlastné spracovanie)

Článok bol prezentovaný na konferencii Nové trendy v marketingu 2012. Vydanie článku bolo v rámci inštitucionálneho projektu FPPV-03-2012 „Aplikácia dolovania dát do metodológie Six Sigma“, riešeného na FMK UCM v Trnave.

Zoznam bibliografických odkazov

- (1) GIUDICI, P. – FIGINI, S.: Applied Data Mining for Business and Industry. Second Edition. John Wiley & Sons Ltd, 2009, ISBN 978-0-470-05886-2
- (2) ČTK: Slováci si obľúbili vernostné karty obchodov, máte ich aj vy? [online]. Bratislava : Ringier Axel Springer Slovakia, a.s., 26.9.2010. [cit. 2012-10-29]. ISSN neuvedené. Dostupné na: <http://www.cas.sk/clanok/181092/slovaci-si-oblubili-vernostne-karty-obchodov-mate-ich-aj-vy.html>
- (3) SITA: Vernostné karty vlastnia vyše štyri pätiny Slovákov. [online]. Bratislava : Petit Press, a.s., 9.9.2012. [cit. 2012-10-29]. ISSN 1335-4418. Dostupné na: <http://ekonomika.sme.sk/c/6526771/vernostne-karty-vlastnia-vyse-styri-patiny-slovakov.html>
- (4) TNS Slovakia: Priemerne vlastnime tri vernostné karty. [online]. Bratislava : Ecopress a.s., 7.9.2012. [cit. 2012-10-29]. ISSN neuvedené. Dostupné na: <http://strategie.hnonline.sk/spravvy/marketing/priemerne-vlastnime-tri-vernostne-karty>
- (5) VAŽAN, P. – KEBÍSEK, M. – TANUŠKA, P. – JUROVATÁ, D.: The data warehouse suggestion for production system. In: Annals of DAAAM and Proceedings of DAAAM Symposium. 2011, Vol. 22, No. 1, p. 0017-0018, ISSN 1726-9679
- (6) HALENÁR, R.: Matlab possibilities for real time ETL method. In: Acta Technica Corviniensis : Bulletin of Engineering, 2012, Roč. 5, č. 1 (2012), s. 51-53, ISSN 2067-3809

Adresa autora:

Andrej Trnka, Ing. PhD.
Univerzita sv. Cyrila a Metoda v Trnave
Fakulta masmediálnej komunikácie
Nám. J. Herdu 2
917 01 Trnava
andrej.trnka@ucm.sk