

DATA MINING V PRAXI: SEGMENTACE ZÁKAZNÍKŮ DLE NÁKUPNÍHO CHOVÁNÍ

DATA MINING IN PRACTICE: CUSTOMER SEGMENTATION BY PURCHASING BEHAVIOR

Julie Poláčková

***Abstract:** The paper focuses on the usage of data-mining techniques as a support tool for decision making. This paper describes the mining of hidden and potentially useful information from databases using data mining methods. These methods, sometimes called as techniques for knowledge discovering, help users, mostly managers, to make qualified decisions in the organization. The aim of the process of knowledge management is not only to collect information, but to transform it into knowledge and use it in a decision-making process. The purpose of this paper was to find and evaluate the different methodological approaches appropriate for customer segmentation. Various data mining techniques were used for demonstration of customer segmentation according to their purchasing behavior within a selected hypermarket. The following techniques were used for clustering: K-means clustering method, Two Step clustering method and Self Organizing Maps. The quality of final models was evaluated by Silhouette measure. It combines the principles of clusters separation and cohesion. Data mining model was constructed from approximately 60 thousand transaction records. Only the food records were selected for the analysis. The paper also examined the effect of the number of dimensions to the clustering. The original variables were reduced into a smaller number of uncorrelated principal components. These components were used for construction of a scatter plot to check the homogeneity of clusters. The results of this analysis confirmed that the reduction of dimensionality is an useful device for the evaluation of generated clusters.*

***Keywords:** Data Mining, Cluster Analysis, Principal Component Analysis, Customer Segmentation, Knowledge Discovering.*

***JEL Classification:** C38, D83.*

1 Úvod

V současném vysoce konkurenčním prostředí je pro organizace, bez ohledu na jejich velikost, velice důležitá schopnost dobře poznat potřeby, očekávání a zejména chování svých zákazníků či odhalení slabých míst podnikových procesů. K dosažení tohoto poznání se dnes stále častěji využívají data-miningové přístupy. Data-mining je vnímán jako soustava predikčních postupů k transformaci datových zdrojů na informaci podporující v organizaci řízení, rozhodování či plnění obchodních cílů. Nejde přitom o nástroj určený pouze pro velké firmy vybavené rozsáhlou technologickou infrastrukturou, podnikovými informačními systémy či datovými

sklady. I ve střední a malé organizaci lze tyto postupy využít s relativně stejným přínosem jako ve velkých firmách. (Vlach, 2006)

Kotler (2007) ve své nejznámější knize Marketing Management upozorňuje na skutečnost, že mezi daty, informacemi, znalostmi a moudrostí existují obrovské rozdíly. Pokud data nejsou zpracována v informace, které se transformují ve znalosti a později přemění v tržní moudrost, značná část jich přijde nazmar. Efektivním a maximálním využitím dostupných dat v organizaci se zabývá tzv. Knowledge Management neboli řízení znalostí. Pokrývá zejména ty procesy v organizaci, které jsou synergií možností informačních technologií při zpracování dat a informací s tvůrčí a inovativní schopností lidských jedinců. V rámci procesu objevování znalostí v databázích je využívána statistická analýza dat, induktivní učení, data mining, genetické algoritmy apod.

Také Drucker (2000) upozorňuje na příhod nové informační revoluce, která se již primárně nesoustředí na shromažďování, ukládání, přenos a prezentaci dat, ale soustředí se na informace, na jejich význam a smysl. Dnešní společnost lze obecně charakterizovat jako společnost, kde znalosti jsou jak pro jednotlivce, tak i pro ekonomiku jako celek primárním zdrojem. Jsou to právě informace, které pracovníkům disponujícím znalostmi umožňují vykonávat svěřenou práci.

Za komplexní poznání je považována množina znalostí, informací a dat vztahujících se k určité problematice. Proces řízení znalostí se stará o to, aby znalosti nebyly pouze hromaděny, ale aby byly racionálně využívány. K tomu patří především šíření znalostí a soustavná cílevědomá analýza účinnosti všech opatření souvisejících s řízením znalostí. Carneiro (2000) uvádí, že vnímání stávajících vztahů mezi objekty zkoumání vede k novým interpretacím a posunuje tak znalosti na vyšší úroveň. Možnosti inovace tedy závisí na znalostech vývoje.

Chiu - Tavella (2008) uvádějí, že efektivní využití data-miningových technik pomáhá identifikovat potenciální příležitosti k realizaci zisků z prodeje, což může vést k vyšší návratnosti investic a tím k vytvoření konkurenční výhody. Pomocí prediktivního modelování je možné získat užitečné informace a lépe pochopit danou problematiku, což dopomáhá k tvorbě informovanějších obchodních rozhodnutí a doporučení.

Berry - Linoff (2004) definují dolování dat jako proces výběru, prohledávání a modelování ve velkých objemech dat sloužící k odhalení dříve neznámých vztahů mezi daty za účelem získání obchodní výhody. Důležitým faktorem úspěšného nasazení metod dolování dat je stručná a srozumitelná prezentace výsledků ve formě přímo použitelné pro rozhodování. Jednou z častých oblastí aplikací je nalezení modelů, které jsou vhodné pro predikci budoucích hodnot atributů na základě nalezených vzorů v datech. Predikce tedy představuje odhadování současných či budoucích hodnot, které nejsou vzhledem ke složitosti a komplexnosti pozorovaného objektu běžnými způsoby odhalitelné či měřitelné.

Rahman (2008) charakterizuje dolování dat jako proces extrakce inteligentní informace z velkého množství surových dat. Kolektivní úsilí při strojovém učení, umělé inteligence, statistické, databázové komunity využívají technologie získávání

znalostí v databázích k získání cenných informací z obrovského množství dat pro podporu inteligentního rozhodování. Data mining se zaměřuje na vývoj algoritmů pro získávání nových vzorů ze skutečných záznamů v databázi. Na otázku proč využívat data-miningové technologie reagují Swingler a Cairns (2008): "...výpočetní inteligence nabízí nové příležitosti podnikům, které si přejí zlepšit efektivitu svého provozu". Data-miningové technologie poskytují pohled do budoucnosti, odpovídají na otázky jako: "Co budou zákazníci kupovat?" nebo "Jaký růst poptávky vyvolá následující reklamní kampaň?"

Cílem předkládaného příspěvku bude nalezení a zhodnocení různých metodických přístupů vhodných k segmentaci zákazníků. Využití adekvátních dataminingových technik bude demonstrováno na konkrétním příkladu seskupování zákazníků zvoleného hypermarketu. Pomocí metod shlukové analýzy bude provedena segmentace zákazníků dle jejich nákupního chování. Dataminingový model bude sestaven na základě cca 60 tis. transakčních záznamů. Analyzována budou pouze data týkající se potravin.

V rámci příspěvku bude dále hodnocena kvalita nalezených řešení pomocí Silhouetovy míry a grafického znázornění získaných shluků. Pomocí nástroje na redukci dimenzionality – analýzy hlavních komponent – budou vstupní proměnné transformovány do menšího počtu vzájemně nekorelovaných hlavních komponent. Tyto komponenty budou následně využity pro grafické znázornění homogenity nalezených segmentů.

Úprava vstupní datové matice, explorační analýza zvolených proměnných, restrukturalizace dat i modelování bude realizováno pomocí dataminingového nástroje Modeler 14, společnosti IBM SPSS.

2 Postup dataminingového modelování

V rámci dataminingového procesu je kladen velký důraz na porozumění analyzovaným datům. Nejprve je proto vhodné provést explorační analýzu sledovaných proměnných a jejich grafické znázornění. Modelovací fázi také předchází úprava vstupních proměnných. Kromě standardizace je možné zhodnotit možnosti logaritmické transformace vstupních proměnných, případně provést redukci dimenzionality ve velkých datových souborech.

Velkou váhu lze přiřadit volbě vhodné dataminingové techniky, k získání konečného řešení je nezbytné zvolit nejvhodnější model z různých dostupných metod. Ke zvolení nejlepšího modelu se využívají odlišná hodnotící kritéria. Každá metoda má potenciál k vyzdvihnutí určitých aspektů, které mohou být jinou metodou ignorovány.

Pro nalezení segmentů zákazníků s obdobným nákupním chováním, lze využít shlukovou analýzu, jejímž cílem je nalezení optimálního seskupení dat, kdy jednotlivá pozorování nebo objekty každého shluku jsou vzájemně podobné, ale jednotlivé shluky co nejvíce rozdílné. Shlukovací techniky patří mezi nástroje nepřímého objevování znalostí. V případě několika málo (2-3) dimenzí lze shluky rozpoznat vlastním okem, s růstem dimenzí se však zvyšuje náročnost vizuálně shluky rozeznat.

Čím vyšší je počet dimenzí, tím více roste důležitost geometrických analýz (Berry – Linoff, 2004). Důvodem k realizaci shlukové analýzy je podle Renchera (2002) předpoklad, že ve zkoumaných datových souborech nalezneme smysluplná přirozená seskupení.

Mezi běžné typy metod shlukování patří hierarchické shlukování a metoda rozkladu (partitioning). Hierarchické shlukování začíná s n shluky, kdy každé pozorování tvoří samostatný shluk, a končí jedním shlukem, který zahrnuje všechna pozorování. Principem metody rozkladu je rozdělení jednotlivých objektů do předem definovaného počtu k shluků. Nejznámější nehierarchickou metodou je metoda k -průměrů (K-means clustering). Jedná se o iterační algoritmus, který minimalizuje součet vzdáleností každého objektu od těžiště shluku. Cílem je získat množinu shluků, které jsou kompaktní a navzájem dobře separované. Dalším využívaným přístupem je metoda dvoustupňového shlukování (Two Step clustering). Tato technika v prvním kroku rozdělí jednotlivých objekty do velkého počtu malých shluků. V druhém kroku jsou pak podobné shluky slučovány tak dlouho, dokud není dosaženo požadovaného množství výsledných k shluků. Mezi moderní techniky shlukování se řadí nesupervizovaný algoritmus založený na principu neuronových sítí zvaný Kohonenovy mapy. Více informací lze nalézt např. v: (Ding, 2004, Lletí, 2004, Rencher, 2002, Řezanková, 2007).

Pomocí těchto seskupovacích technik a jejich různých možných nastavení lze získat velké množství výsledných modelů. Je proto důležité zvolit vhodné kritérium k porovnání získaných řešení (Bae, 2010). K hodnocení kvality nalezených řešení lze využít např. Silhouetovou míru, která kombinuje principy kohenze a separace shluků.

V rámci komplexní analýzy lze následně sledovat grafické znázornění získaných seskupení. Pomocí nástroje na redukci dimenzionality – analýzy hlavních komponent – lze vstupní proměnné transformovat do menšího počtu vzájemně nekorelovaných hlavních komponent. Tyto komponenty mohou být graficky znázorněny v trojrozměrném bodovém grafu, kde každý bod bude obarven dle příslušnosti k definovanému shluku.

Hlavními cíli analýzy hlavních komponent jsou na jedné straně nalezení správného rozměru souboru dat a tím bez výrazné ztráty informace zlepšení kvality analýzy a na straně druhé nalezení nových proměnných. Jedná se v podstatě o transformaci původních proměnných x_i , $i = 1, \dots, m$, do menšího počtu latentních proměnných y_j . Tyto proměnné mají vhodnější vlastnosti, jejich počet je výrazně nižší, vystihují téměř celou proměnlivost původních proměnných a jsou vzájemně nekorelované. Latentní proměnné jsou u této metody nazvány hlavními komponentami a jde o lineární kombinace původních proměnných: první hlavní komponenta y_1 popisuje největší část proměnlivosti čili rozptylu původních dat, druhá hlavní komponenta y_2 zase největší část rozptylu neobsaženého v y_1 atd. Podrobnější informace lze nalézt např. v: (Field, 2005, Hebák Et Al., 2007, Lavine, 2000, Meloun A Militký, 2001).

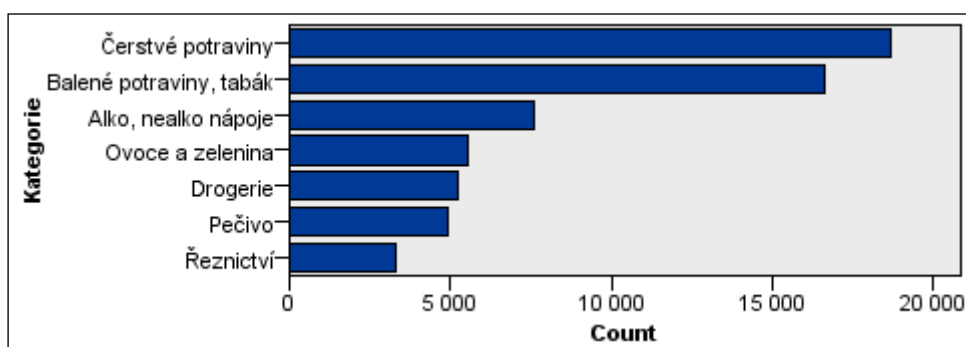
3 Segmentace zákazníků dle nákupního chování

3.1 Explorační analýza

Nejprve byla provedena explorační analýza jednotlivých vstupních proměnných. V přípravné fázi bylo také důležité zvážit možnosti restrukturalizace datové matice. Konkrétně byla provedena restrukturalizace jednotlivých transakcí tak, aby každý řádek odpovídal jednomu nákupnímu košíku. Modelovací fázi také předchází úprava vstupních proměnných. Byla provedena standardizace proměnných, aby nedocházelo ke zkreslení výsledků shlukové analýzy. Dále byla zvažována logaritmická transformace, která však nepřinesla výrazné zkvalitnění výstupního modelu. Aby nedošlo k ovlivnění výsledné analýzy vybočujícími hodnotami, byly hodnoty přesahující trojnásobek směrodatné odchylky upraveny pomocí metody winsorizace, tedy nahrazením odlehlých hodnot hodnotou předcházející.

V této fázi projektu byla navržena vhodná struktura datového souboru ve formě datové matice. Součástí bylo zvolení proměnných vstupujících do modelu: hodnota nákupního košíku v Kč, počet položek v košíku, podíl ceny akčního zboží a kategorie produktů obsažených v košíku. Proměnná podíl akčního zboží byla vyčíslena jako podíl celkové ceny akčního zboží v košíku a hodnoty celého nákupního koše v Kč.

Následující obrázek (obr. 1) zachycuje rozdělení vstupního souboru do jednotlivých kategorií potravin.



Obr. 6: Zastoupení položek v jednotlivých kategoriích

Zdroj: Autor

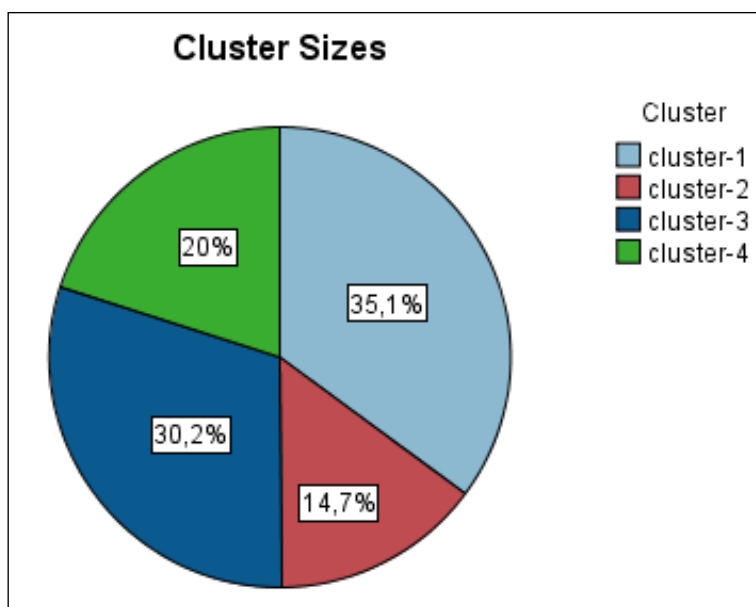
K modelování byly zvoleny právě tři vstupní kategorie zboží, jelikož model se třemi kategoriemi vykázal lepší segmentační kvality než modely s větším počtem kategorií. Konkrétně byly zvoleny tyto kategorie potravin: první kategorie (FOOD1) obsahovala alkoholické a nealkoholické nápoje, balené potraviny a cigarety a drogerii, druhá kategorie (FOOD2) zahrnovala čerstvé potraviny (např. balené uzeniny a masné výrobky, mléčné výrobky apod.), pečivo a ovoce a zeleninu, poslední kategorie představovala produkty řeznictví (FOOD3).

3.2 Fáze modelování

V modelovací fázi byla provedena segmentace zákazníků. Cílem segmentační analýzy je nalezení takového seskupení, kdy objekty každého shluku jsou si vzájemně co nejvíce podobné, ale samotné shluky navzájem co nejvíce rozdílné. Tyto techniky nejsou z pravidla využívány samostatně, nýbrž ve spojení s dalšími metodami. Jakmile jsou identifikovány příslušné segmenty, je aplikována další metoda, pomocí které se zjistí význam jednotlivých segmentů.

K samotné realizaci segmentace nákupních košů byly využity následující techniky shlukování: metoda k-průměrů, metoda dvoustupňového shlukování a Kohonenovy mapy. K hodnocení kvality nalezených řešení byla využita Silhouetova míra, která kombinuje principy kohenze a separace shluků. Nejlepší modely vytvořila metoda k-průměrů se 4 či 5 výslednými shluky. Jedná se o jednu z nejčastěji využívaných segmentačních technik. Oba tyto modely dosáhly hodnoty 0,51 Silhouetovy míry, což představuje dostatečně kvalitní model.

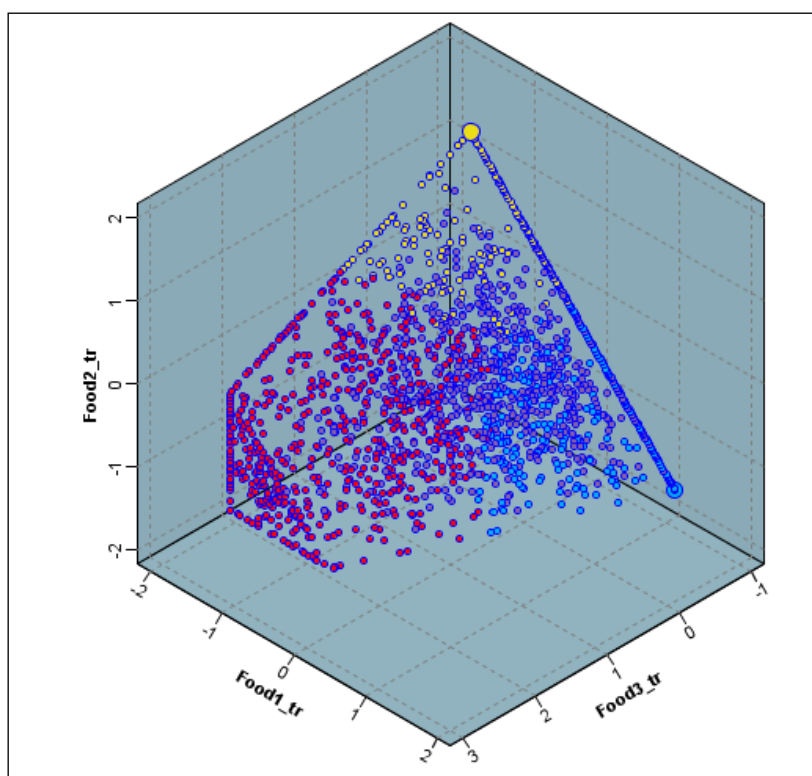
Další vhodný model byl získán pomocí metody dvoustupňového shlukování. Tento model s počtem 8 shluků vykázal pouze o setinu nižší hodnotu Silhouetovy míry (0,50). Dále byly zkoumány možnosti shlukování pomocí nesupervizovaného algoritmu založeného na principu neuronových sítí zvaného Kohonenovy mapy. Tato technika dosahovala nízké úrovně Silhouetovy míry a poskytla velké množství malých shluků (20). Jako nejvhodnější byl zvolen model se 4 shluky získaný metodou k-průměrů, zachycený na obrázku č. 2. Tento model poskytl homogenní řešení a logicky odůvodnitelné segmenty. Jako nejdůležitější proměnné pro modelování algoritmus zvolil všechny vstupní proměnné s výjimkou podílu akčního zboží.



Obr. 7: Výsledné shluky získané pomocí metody K-means

Zdroj: Autor

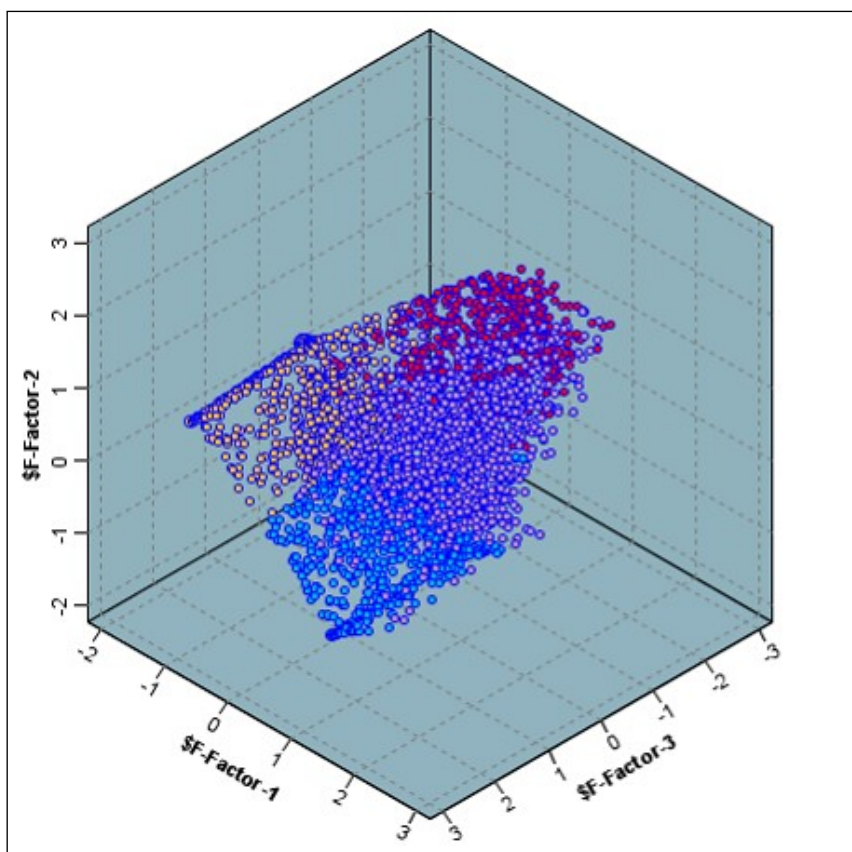
Na následujícím diagramu (obr. 3) je možné pozorovat rozdělení jednotlivých nákupních košů do získaných segmentů dle zvolených proměnných. Tento trojrozměrný diagram byl získán zachycením tří zvolených proměnných - kategorie FOOD1 (nápoje, tabák, drogerie), FOOD2 (čerstvé potraviny), FOOD3 (řeznictví) – do bodového grafu. Obdobným způsobem byly hodnoceny grafy ostatních proměnných. Cílem grafického znázornění je zhodnotit homogenitu získaných shluků. Hodnocení dostatečné separovanosti shluků je v tomto případě obtížné, jelikož vždy byly zachyceny pouze tři z celkového počtu šesti analyzovaných proměnných. Z tohoto důvodu shluky nepůsobí zcela homogenním dojmem, i tak je ale patrná jejich oddělitelnost.



Obr. 8: Bodový graf tří hlavních kategorií potravin obarvených dle příslušnosti ke shluku

Zdroj: Autor

Dále byla zkoumána možnost redukce dimenzionality datového souboru, což by mohlo přispět k výraznější separaci výsledných shluků v grafu. Z tohoto důvodu byla provedena analýza hlavních komponent, pomocí níž byly vstupní proměnné transformovány do tří navzájem nekorelovaných latentních proměnných. Tyto nové proměnné vysvětlují 83,3 % celkové variability modelu. Trojrozměrný graf získaný pomocí prvních tří hlavních komponent zachycuje homogenní skupiny shluků. Výstupy z této analýzy potvrdily, že snížení dimenzionality usnadní hodnocení kvality získaného řešení.



Obr. 9: Bodový graf tří hlavních komponent obarvených dle příslušnosti ke shluku

Zdroj: Autor

3.3 Charakteristika získaných segmentů

Segmentace rozdělila skupinu všech zákazníků do menších homogenních celků, které se vzájemně liší potřebami zákazníků, jejich charakteristikami a nákupním chováním. Následná definice segmentů usnadní řídicím a marketingovým pracovníkům oslovovat každý segment odlišnými nabídkami, které budou směřovány právě na potřeby zákazníků v daném segmentu.

Finálním krokem segmentace je tedy identifikace a podrobnější analýza jednotlivých nalezených segmentů. První a zároveň největší segment zahrnuje více než třetinu všech zákazníků (35 %). Tento segment lze charakterizovat vysokým podílem zboží z kategorie FOOD1 (nápoje, drogerie, balené potraviny a cigarety), naopak nízkým podílem zbývajících kategorií FOOD2 (čerstvé potraviny, pečivo a ovoce a zelenina) a FOOD3 (řeznictví). Jedná se tedy o relativně malé nákupní košíky s nižší celkovou hodnotou, které obsahují především nápoje, drogerii, balené potraviny a cigarety. Průměrná hodnota těchto košíků je 222,- Kč, průměrný počet položek je 6 až 7. Podíl akčního zboží v tomto koši činí cca 20 %.

Druhý segment zahrnuje nejnižší podíl zákazníků (15 %). V tomto segmentu výrazně převažuje kategorie FOOD3, nižší podíl zastává kategorie FOOD2, naopak FOOD1 se v těchto koších téměř nevyskytuje. Jedná se o středně velké koše průměrné

hodnoty obsahující především produkty z řeznictví a čerstvé potraviny. Průměrná cena těchto košů je 358,- Kč, průměrný počet položek v košíku je 11. Tento segment obsahuje nejnižší podíl akčního zboží, v průměru pouze 13 %.

Třetí a zároveň dle podílu zákazníků (30 %) druhý největší segment lze charakterizovat vysokým podílem kategorie FOOD2. Za toto zboží zaplatí zákazníci v průměru 83,- Kč. Naopak nízký podíl zastává kategorie FOOD1. Do tohoto segmentu spadají nejmenší nákupní košíky s průměrným počtem položek pět a s nejnižší průměrnou celkovou hodnotou 111,- Kč. Tyto koše obsahují především čerstvé potraviny, pečivo či ovoce a zeleninu.

Poslední čtvrtý segment, do kterého patří cca pětina celkových zákazníků, lze charakterizovat především vyváženým obsahem. Mírně převažuje kategorie FOOD1 (nápoje, drogerii, balené potraviny nebo cigarety) a FOOD3 (řeznictví). S průměrným počtem položek 36 se jedná o segment největších a nejdražších nákupních košů, s průměrnou celkovou hodnotou 1150 Kč.

Samotná segmentace z pravidla není finálním krokem analýzy nákupního chování, získané shluky dále vstupují jako vstupní proměnné do následných analýz, např. do asociační analýzy nákupních košíků. Tento přístup je vhodný především v případě předpokladu, že jednotliví zákazníci se budou v rámci různých segmentů chovat odlišně.

4 Závěr

Předkládaný příspěvek má především metodický charakter. V rámci příspěvku byly zhodnoceny různé přístupy k segmentaci zákazníků dle jejich nákupního chování. K realizaci segmentace byly využity následující techniky shlukování: metoda k-průměrů, metoda dvoustupňového shlukování a Kohonenovy mapy. Kvalita nalezených řešení byla hodnocena pomocí Silhouetovy míry. Nejlepší modely vytvořila metoda k-průměrů se čtyřmi výslednými shluky.

Homogenita výsledných shluků byla hodnocena pomocí trojrozměrných diagramů, do kterých byly vyneseny tři zvolené vstupní proměnné. Jednotlivé nákupní koše byly obarveny dle příslušnosti k vytvořenému segmentu. Hodnocení dostatečné homogenity shluků bylo v tomto případě obtížné, jelikož vždy byly zachyceny pouze tři z celkového počtu šesti analyzovaných proměnných. Z tohoto důvodu byla zkoumána možnost redukce dimenzionality datového souboru, což přispělo k zachycení velkého podílu variability původních vstupních proměnných. Trojrozměrný graf, získaný pomocí prvních tří hlavních komponent, již zachytil dobře separované a homogenní shluky. Potvrdila se tak kvalita nalezeného řešení a vhodnost využití tohoto přístupu při hodnocení výstupů shlukové analýzy.

Cílem příspěvku bylo nalezení a zhodnocení různých metodických přístupů vhodných k segmentaci zákazníků. Bylo prokázáno, že metoda k-průměrů v kombinaci s analýzou hlavních komponent představuje vhodnou techniku segmentace zákazníků dle nákupního chování. Výsledná segmentace rozdělila skupinu všech zákazníků do menších homogenních celků, které se vzájemně liší potřebami zákazníků a jejich

nákupním chováním. Výsledné segmenty tedy zachycují existující rozdíly mezi jednotlivými skupinami zákazníků.

V následujícím výzkumu budou tyto segmenty využity pro podrobnější vyhodnocení nákupního chování zákazníků pomocí asociačních pravidel mezi jednotlivými typy produktů. Samotná definice segmentů společně s dalšími poznatky usnadní marketingovým pracovníkům účelnější a efektivnější komunikaci směrem k zákazníkovi. Ve finálním důsledku může vést segmentace k získání konkurenční výhody. Zvyšování konkurenceschopnosti totiž vyžaduje rozhodovací procesy postavené na spolehlivých a vhodných informacích, získaných z dostupných interních a externích zdrojů.

Poděkování

Tento článek byl zpracován s podporou výzkumného projektu: IGA, PEF ČZU, č. 201011170031, „Aplikace data-miningových technik v oblasti prediktivního modelování“.

Použité zdroje

- [1] BAE, E., BAILEY, J., DONG, G. A clustering comparison measure using density profiles and its application to the discovery of alternate clusterings. *Data Mining and Knowledge Discovery*, 2010, Vol. 21, No. 3, pp. 427-471. ISSN: 1384-5810.
- [2] BERRY, M., LINOFF, G. *Data Mining Techniques For Marketing, Sales, and Customer Relationship Management*. Second Edition. Indianapolis: Wiley Publishing, 2004. 643 p. ISBN 0-471-47064-3.
- [3] CARNEIRO, A. How does knowledge management influence innovation? *Journal of Knowledge Management*, 2000, Vol. 4, No. 2, pp. 87-98. ISSN 1367-3270.
- [4] CHIU, S., TAVELLA, D. *Data Mining and Market Intelligence for Optimal Marketing Returns*. First Edition. Oxford: Elsevier Inc, 2008. 296 p. ISBN: 978-0-7506-8234-3.
- [5] DING, C., HE, X. K-means Clustering via Principal Component Analysis. *Proceedings of the 21st International Conference on Machine Learning*, New York, 2004. ISBN:1-58113-838-5.
- [6] DRUCKER, P., F. *Výzvy managementu pro 21. století*. Praha: Management Press, 2000. 188 s. ISBN: 80-7261-021-X.
- [7] FIELD, A. *Discovering Statistics Using SPSS*. London: SAGE Publications, 2005. 816 p. ISBN: 0-7619-4451-6.
- [8] HEBÁK, P. a kol. *Vícerozměrné statistické metody [3]*. Praha: Informatorium, 2007. 255 s. ISBN 978-80-7333-001-9.
- [9] KOTLER, P. *Marketing Management*. 12. vydání. Praha: Grada, 2007. 792 s. ISBN: 80-247-1359-4.

- [10] LAVINE, K. B. Clustering and Classification of Analytical Data. *Encyclopedia of Analytical Chemistry*, Chichester: John Wiley & Sons Ltd, 2000. p. 9689-9710. ISBN: 9780470027318.
- [11] LLETÍ, R., ORTIZ, M.C., SARABIA, L.A., SÁNCHEZ, M.S. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Analytica Chimica Acta*, 2004, Vol. 515, No. 1, pp. 87-100, ISSN 0003-2670.
- [12] MELOUN, M., MILITKÝ, J., HILL, M. *Počítačová analýza vícerozměrných dat v příkladech*. Praha: Academia, 2005. 449 s. ISBN 80-200-1335-0.
- [13] RAHMAN, H. *Data Mining Applications for Empowering Knowledge Societies*. Bangladesh: Idea Group Inc., 2008. 356 p. ISBN 978-1599046570.
- [14] RENCHER, A. *Methods of Multivariate Analysis*. Second Edition. New York: Wiley Publishing, 2002. 738 p. ISBN 0-471-41889-7.
- [15] ŘEZANKOVÁ, H. *Analýza dat z dotazníkových šetření*. Příbram: Professional Publishing, 2007. 217 s. ISBN 978-80-86946-49-8.
- [16] SWINGLER, K., CAIRNS, D. *Making Decisions with Data: Using Computational Intelligence Within a Business Environment*. Bangladesh: Idea Group Inc., 2008. 356 pp. ISBN 978-1599046570.
- [17] VLACH, P. Data mining v malých a středních organizacích. *Small Business Solutions*, 2006, č. 12, s. 20-22. ISSN 1802-615.

Kontaktní adresa

Ing. Julie Poláčková

Česká zemědělská univerzita

Provozně ekonomická fakulta, Katedra statistiky

Kamýcká 129, 165 21 Praha 6 – Suchbátka, Česká republika

Email: polackova@pef.czu.cz

Tel.: +420 22438 2299

Doručeno redakci: 29. 04. 2011

Recenzováno: 15. 07. 2011

Schváleno k publikaci: 09. 08. 2011