An Approach to Discovery of Customer Profiles

Ilona Pawełoszek[1(✉)] and Jerzy Korczak[2]

[1] Częstochowa University of Technology, Częstochowa, Poland
ipaweloszek@zim.pcz.pl
[2] Wrocław University of Economics, Wrocław, Poland
jerzy.korczak@ue.wroc.pl

Abstract. The goal of the paper is to present the opportunity of exploiting data analysis methods and semantic models to discover customer profiles from financial databases. The solution to the problem is illustrated by the example of credit cards promotion strategy on the basis of historical data coming from the bank's databases. The database contains information, personal data, and transactions. The idea is founded on data exploration methods and sematic models. With this purpose in mind, multiple algorithms of clustering and classification were applied, the results of which were exploited to elaborate the ontology and to define the customer profile to be used in decision-making.

Keywords: Customer profile · Data mining · Ontology of marketing · Semantic models

1 Introduction

In the age of personalization, one of the greatest challenges for marketers is eliciting and communicating customer requirements. A customer profile — also known as a customer persona — is a set of data describing a high-level abstraction model that depicts the key characteristics of a group of consumers who could be interested in a specific product. Personas are fictitious, specific, concrete representations of target users [1].

Construction of customer profiles promotes overall internal alignment and coordination of marketing strategy with product development which helps to drive down cost of promotion by reducing the number of useless marketing messages and ineffective contacts with customer. Although making use of customer feedback is an established method of gathering marketing intelligence, interpreting data obtained by traditional structured methods such as questionnaires, interviews and observation, is often too complex or too cumbersome to apply in practice [1, 2].

Today, a lot of data can be gathered in an automatic way from transactional systems. These data describe the features of customers (such as age, the place of residence, number of family members, etc.) as well as behaviors (such as time and purpose of transactions performed by the customer, amounts and frequency of expenditures).

A.M. Tjoa et al. (Eds.): CONFENIS 2016, LNBIP 268, pp. 88–99, 2016.
DOI: 10.1007/978-3-319-49944-4_7

The discovery of useful knowledge from databases is one of the major functions of analytical decision support systems [3, 4]. Various data mining algorithms can be used to discover and describe behavior patterns of individuals and to relate them with personal data collected in CRM data bases.

Unfortunately, most of the systems lack semantics, which can be considered a key component for business knowledge management. Marketing databases are today very broad. The volume of information is so huge that the analysis by classical database methods is becoming increasingly difficult. In recent years a few solutions have appeared improving the process of database exploration and the integration of semantics in decision-making processes [5–11].

A few proposals for marketing ontologies have appeared in recent years. Barbu [12] presents General Marketing Ontology, in which the main concepts originate from sources such as Web marketing repositories and tools allowing for the semi-automatic creation of ontology on the basis of documents important to managers of SMEs. Saggion describes a project of integration in the BI system, founded on the module of gathering information in an e-business domain using ontology and natural language [13]. The documents describing profiles of companies are collected from different sources, analyzed, grouped and annotated within the ontology. The semantic models are also exploited in many search engines dedicated to marketing information, i.a. Magpie, KIM, SemTag, On-To-Knowledge, Vision, h-TechSight [14–16]. The methods of data exploration have become a very popular tool used in marketing strategies [17–20]. It is a well-known fact that the effectiveness of marketing activities depends to a large degree on directing relevant advertisements to the right recipients. One of the essential functions of customer relationship management is customer segmentation which is the process of dividing the customer data set into distinct and internally homogeneous groups in order to develop differentiated marketing strategies according to their characteristics [21]. With this aim in mind, a segmentation of customers consists in finding customers with similar preferences, needs and behavior. A segment corresponds to target class of customers, for example, travelers that are people with high income who travel often. In case of large data sets the experiential classification can be insufficient to distinguish segments. In such situations data mining methods such as clustering and decision trees can be helpful. The terms of *cluster* and *segment* are often used interchangeably, which is not always the same.

Clusters are groups of customers with similarities found by a clustering algorithm. Therefore, a cluster not necessarily represents a segment, so the semantic interpretation of clusters is required and might be difficult. It should be noted that in many cases the segment can consist of a few clusters.

The segmentation of customers not only facilitates choosing a right product, but also allows one to provide the customers with information that would be interesting for them. The new knowledge obtained from data exploration processes and presented as ontology on one hand describes concepts and relations, and on the other facilitates access to the important and useful information in the marketing database for a manager [22].

Information on the place of the customer's residence is currently not enough for product promotion purposes. The company should know also who the customer is, what his or her characteristics are, what is valuable, what is attractive, and what is completely uninteresting or even annoying. For example, to target the advertising to young people, it is better to prepare a marketing campaign on the Internet, but the internet advertising may not be so effective in the case of a product targeted to older people, who usually are not familiar with computers.

Marketing applications increasingly often use customers' psychographic profiles [23, 24], developed on the basis of information which is continuously collected during interactions between the company and its customers. The companies seek to register every possible communication record of the customer (phone calls, logging into the system, loyalty cards, using mobile applications, payments, history of transactions, etc.). Determining the customer's needs in advance, and an individual approach to each of them, is becoming the aim of marketing activities of large organizations.

The structure of this paper is as follows. In Sect. 2, the problem of customer profile discovery is defined. In the next section, the sources of data necessary to build classification and prediction models are presented, with particular emphasis on clustering algorithms and classification trees. The last section presents the case study of data exploration methods along with semantic models and their role in developing the customers' profiles in the context of promotion of payment cards.

2 The Problem of Developing the Customer's Profile

Many approaches to building a marketing strategy are described in the literature [17, 19, 25–27]. In this paper, in accordance with marketing theory literature, the creation of customers' profiles is emphasized as an essential tool of marketing strategy implementation [28].

The focus of the project was put on discovery of the bank customer's profile with the aim of identifying potential recipients of dedicated payment cards. Generally the main source of information in the research on a customer's profile is the history of transactions on the customer's bank account. Usually, historical data is complemented by personal information (such as: age, gender, place of residence, number of children) which has undeniable influence on the consumer's preferences and thus may constitute a good base for the customer's classification and for choosing a suitable payment card. This information, its scope and quality, has a significant impact on the predictive quality of the developed model.

Well-targeted promotion is a key determinant of the effectiveness and efficiency of a marketing campaign. The initial choice of a product, made on the basis of information about consumers' preferences, can reduce the time of marketing phone calls, and ensure the right choice of communication channels to reach the potential customers. It reduces ineffective contacts with customers who might not be interested in a particular offer. Finally, better communication influences the effective utilization of resources, which should decrease the costs of promotion.

On the payment cards market there are a large number of products dedicated to various groups of customers, for example: students, people who travel often, seniors, young parents, etc. To make these considerations more precise, let us assume that the bank is going to propose payment cards to its customers, corresponding to the following customer segments:

– "*Travelers"* a dedicated credit card for frequent travelers, offering all kind of additional discounts on airline tickets, accommodation, hotels, insurance.
– "*Eternal students*"- a credit card dedicated to the so-called "eternal students", offering extra discounts on concerts and events.
– "*Still young*" – a credit card dedicated for seniors, offering discounts for coffee bars, additional health insurance, free medical exams.
– "*High heels*" – a credit card dedicated to women, offering additional discounts to shoe stores and clothing.
– "*Business card*" – a card dedicated to business people, offering additional discounts for elegant clothing stores or tickets in Business Class.

The five segments have been briefly characterized by marketing analysts on the basis of their professional experience. Therefore the goal of the project was to improve the scope and specification of segments using data mining methods.

The specificity of the offered cards has decisive influence on the selection of data for analysis. For example, the "*Travelers*" card should be interesting to young or middle-aged people who are financially well-off. Their transaction history is dominated by multiple expenditures for touristic purposes, such as tickets, travel agency services, expenditures and mobile top-ups made from abroad. The target group for "*Eternal students*" cards should be addressed to young people, not having children (their profiles are characterized by i.a. low expenditures on articles for children). Considering the discounts for various events offered, this card can be proposed to people whose transaction history reveals a party lifestyle. The "*Still young*" card should be offered to older people who have significant expenditures on healthcare services. The "*High heels*" card should be interesting mainly for middle aged women. Important information for choosing this type of card will be whether the customer has children and therefore increased expenditures on children's articles and healthcare services. The potential customers for the "*Business card*" are middle-aged people, settled down and having substantial income.

3 Description of Data

In the project, the data about customers was extracted from transaction-oriented systems and mobile banking applications. The experimental data file contains 200 000 anonymized records of customers, describing their personal data, bank products at their disposal, incomes, expenditures and financial transactions The data underwent statistical analysis and initial transformation. More detailed information about the data can be found in [22].

Taking into consideration the selected algorithms of data exploration, the data were converted to numerical values, including continuous, discrete or binary types. The attributes of very low variances and redundant ones were eliminated. In the case of missing values, the descriptions were completed using the nearest neighbor algorithm [20, 29]. On the file prepared in such a way, the normalization was performed which was necessary for heterogeneous variables. As a result of the initial transformations and features reduction, the set of 200 000 observations with 24 variables of the interval [0,1] was given for the further phases of data exploration.

4 Process of Determining Customer Profile

The study assumes that the bank intends to expand its offer for customers of one of five dedicated payment cards. The promotion is intended to avoid the phenomenon of miss-selling, or incorrect identification of the targeted groups of customers.

Generally speaking, the primary idea of the solution was to divide the customer database into semantically interpretable clusters, and then to design a classifier which will permit accurate definition of the customer profiles. In the final task of exploration, the results is used to build the ontology of knowledge about the customers and its use in marketing decision making [17].

For clustering, four algorithms were chosen: k-means, SOM neural networks, hierarchical clustering, and supervised by the user [29, 30]. The obtained partitions by the first three algorithms were very similar. The results showed to some degree not only imperceptible clusters by traditional marketing methods, but also unnoticeable relationships between characteristics of customers and offered products.

Taking as a criterion for aggregation the distance between clusters and semantics of the data, five clusters corresponding to the five potential market segments were

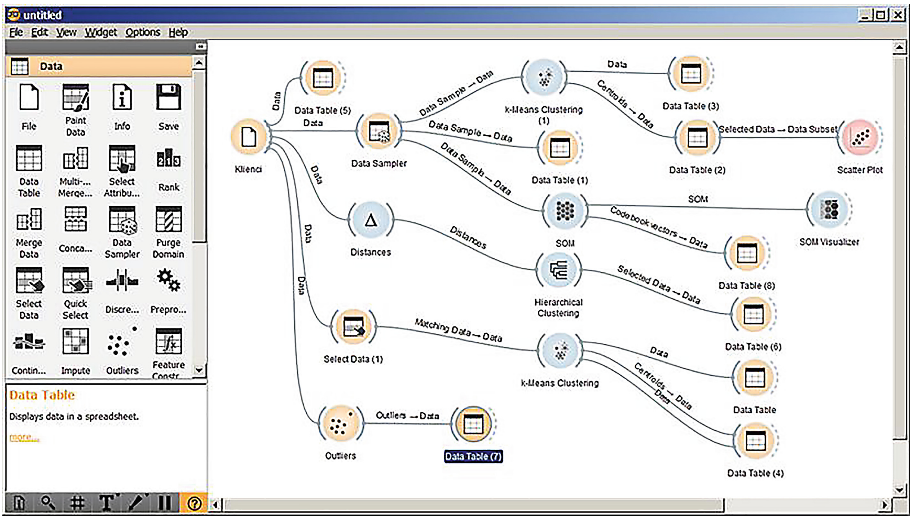

Fig. 1. Diagram of the data mining process

generated. Although it has been verified by specialists, this partition did not show clusters of good quality in the marketing sense. Carrying out further clustering algorithms using k-means clustering and hierarchical clustering showed the groups of customers more homogenous in terms of semantics, but still not satisfactory. Therefore, a new method of clustering was proposed to guide the marketing specialist.

Clustering with user engagement was founded on the preliminary elimination of irrelevant attributes related to the specificity of the clusters sought. This operation made it possible to determine subsets with the high possibility of finding clusters semantically interpretable. For example, one can limit the search space for the clusters associated with "*High heels*" segment by eliminating irrelevant attributes such as: number of transactions on fuel, travelling, healthcare, number of children and mobile top-ups. The most relevant attributes were: gender (woman), age (20 to 60 years old) and monthly income (higher than 1,500 PLN).

Within these constraints the data file containing 200 thousand instances was reduced to approx. 60 thousand clients, women who fulfilled these restrictions. Figure 1 shows a diagram of the process of the data mining platform using Orange[1].

In order to interpret the obtained clusters, four classifiers were applied: inductive decision trees, Multi-Layer Perceptron, Naive Bayes and CN2 [20, 32].

Comparing with the previous publication [22], in this paper the research was extended to all classes of customers.

During the analysis of the clusters, it turned out that women to whom "*High Heels*" can be addressed were dispersed into five clusters C1, C2, C3, C4 and C6. Of these five clusters, the marketing analysts were particularly interested in the clusters with the highest number of instances, C6 and C1, whose initial definitions were extracted from the decision tree (Fig. 2), namely:

C1 ∪ C6: ((Gender true) and (Amount_Entertainment_Transactions <= 25,951) and ((CreditCard true) and (Village True)) or ((CreditCard false) and (Village false)))

In the process of rule validation, the analysts have rejected the condition concerning *Amount_Entertainment_Transactions* and turned it into the condition (*EntertainmentTransactions true*), indicating women who have done at least one entertainment transaction, regardless of the amount.

Legend: The values refer to the nodes of the cluster ID, the degree of homogeneity of the cluster (in %), and the number of instances. The names of the attributes in the nodes have been shortened for visual reasons.

Finally, after all transformations the rules describing the considered classes-were the following:

- Class "*High heels" if ((Gender women) and (EntertainmentTransactions true))* Note that the new class definition is different from the one specified in the suggested restrictions referred to age and income. The final definition of potential customers for a "*High Heels*" card was given in Fig. 4.

[1] Orange is an open source data visualization and data analysis package for data mining applications, developed by Bioinformatics Labs at University of Ljubljana, Slovenia (http://orange.biolab.si) [31].

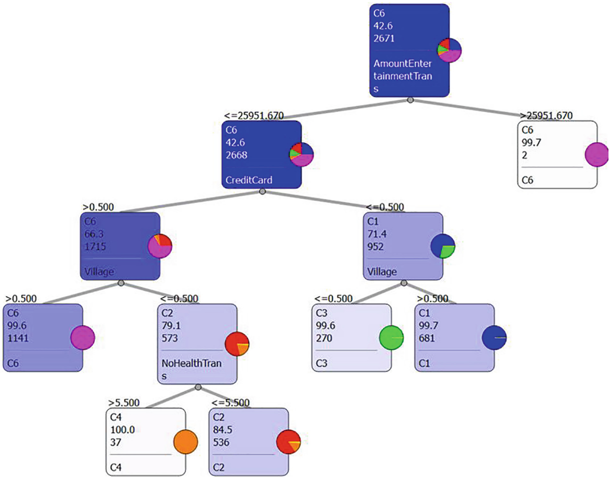

Fig. 2. Fragment of a decision tree for clusters C1 and C6

- Class *"Travellers" if ((AmountRailwayTrans >= 1,000) or (AmountPetrolTrans >= 1,000))*
  The class describes the people who travel a lot, thats why both amounts of railway and petrol transactions are high. The customers use both train and car transport. The new sample of data marke das "*Travellers*" consists of 1,175 persons.
- Class "*Business Card"if (Age <=60 and MonthlyIncome >= 7,000)*
- To choose the right potential customers for *Business Cards* we were searching for women and men with higher than average level of income –as per our assumption it is more than 7,000. Usually they are middle age and older persons but the assumption was made that maximum age is 60 years old. New sample of data is marked as "*Business Cards*" class and it consists of 500 customers. This class is the smallest one, average income is 10,209;
- Class "*Eternal students"- if (Age <=40 and AmountOfTransEntertainment >= 1,000)*
- People who could be interested in the offer "*Eternal Students*" are young people who spent more than 1,000 PLN on entertainment transactions. The number of items in this class is 3,762.
- Class "*Still young" – if (Age >= 55 and AmountOfTransHealth >= 1,000*)
  The class describes older people who spend significant amounts on health transactions. The number of customers who could be interested in this kind of payment card is 1,071.
  After the completion of the data exploration, the construction of the customer ontology started by using the Protégé platform[2]. The schema of part of the ontology

[2] Protégé is a ontology development platform supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health. (http://www-med.stanford.edu).

is shown in Fig. 3. Classes were a priori posed as the groups of customers interested in specific payment cards, indicated in the previous section. Rules and discriminant attributes of the decision tree designated the characteristics of the concepts in the ontology, through the so-called *Data Properties*, and made it possible to define the customer groups (Fig. 4).

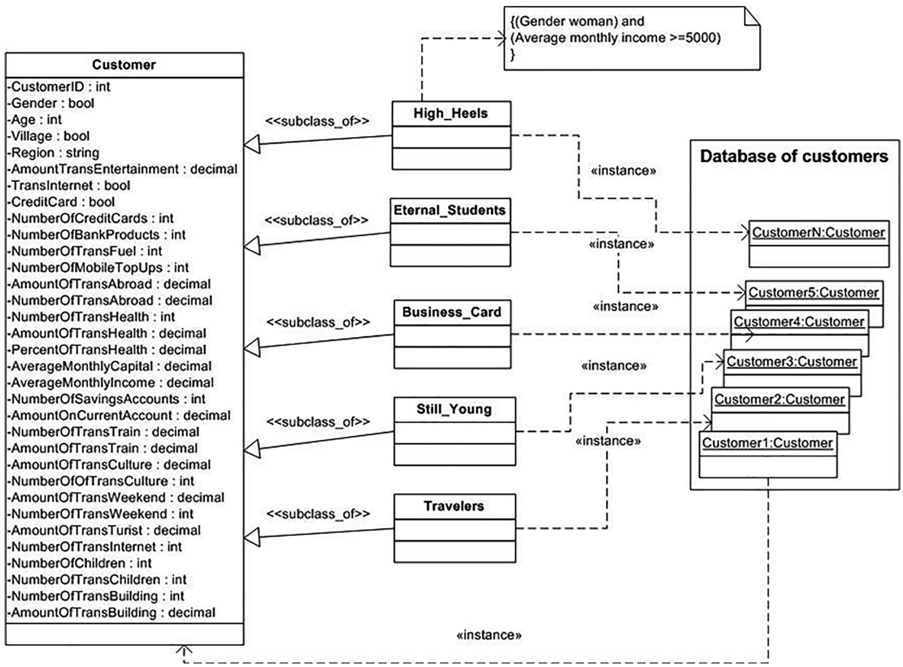

Fig. 3. Diagram of the ontology of customer market segmentation

The ontology so created containing knowledge of the marketing profiles of customers was used not only for information retrieval from the database, but also in the classification of new customers of the bank in connection with the specificity of the products offered.

5 Using the Semantic Model of Customer

The example shows the utility of ontologies in making marketing decisions related to the preparation of the offer for the bank's customers with the explanation of why a particular card should be offered. It was assumed that the manager does not have complete knowledge of business informatics, and, therefore, that she or he does not know SQL programming nor database schema, nor complex functionalities of the data mining platforms. Protégé has been used to define the semantic model in our case study. The financial data was provided in a format of the Microsoft Excel spreadsheet. To enable ontology content collection the Cellfie plugin[3] was used. The Cellfie desktop editor was used to map and to annotate the data in the spreadsheet (columns and rows) to attributes within the financial ontology. The discovered by the data mining algorithms class descriptions were transformed into the set of design patterns. Using knowledge about design patterns improves the database consistency and reduces the amount of time required to validate the data.

The analyst uses only the graphical interface of Protégé, which allows him to consult the ontology of knowledge about customers of the bank (partially shown in Fig. 3), and facilitates access to the marketing database. Figure 4 shows the example of a class description of potential customers for the *"High heels"* card. This diagram illustrates on the left the classes of customer cards, and on the right the definition in the form of logical expression describing this class, and at the bottom a list of customers which meet these requirements. Based on this definition, the manager can not only search the database of all customers who meet the class constraints, but also strengthen or weaken the conditions of the class description. It should be noted that these operations do not require the manager's expertise on database structure and language to formulate queries.

Generally, the data mining methods make it possible to retrieve different clusters or groups of customers and their marketing interpretation. The interpretation, written in

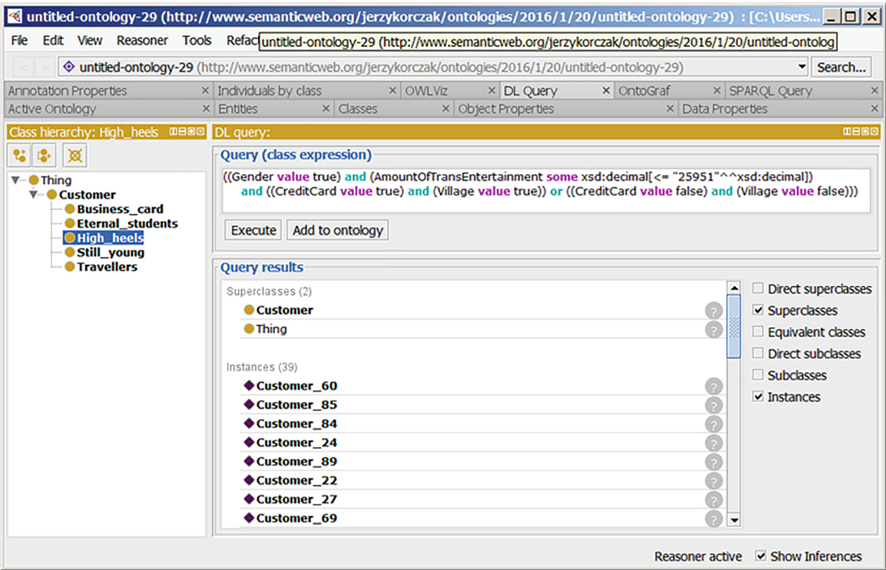

Fig. 4. Definition and instances of the *"High heels"* class

[3] Cellfie is a Protégé Desktop plugin for mapping spreadsheets to OWL ontologies available at https://github.com/protegeproject/cellfie-plugin.

the form of formal expressions in Protégé, enables the manager to access detailed information about the customers, about groups of customers, dependencies, all axioms of marketing knowledge. Concepts, as well as the relationships between them can be easily changed and updated as a result of acquiring new knowledge, new data or new manager experiences.

6 Summary and Directions for Further Research

The paper has presented an analysis of the marketing database containing personal data as well as transactional and financial information. It was shown how one can construct customer profiles through the use of data mining methods. For this purpose, a number of algorithms for clustering and classification were applied. Note that the database schemas are usually designed for efficient data storage, but do not provide a semantic description to facilitate understanding of the data, interpretation, and reasoning. Therefore, the results of data mining were used to construct ontology which describes customer profiles.

The use of ontology with an easy to learn visual interface of the Protégé platform allows managers to learn more about the information contained in the databases; it provides clear definitions related to the attribute names in the database. In addition, the proposed ontology contains the pre-defined classes that automatically make it possible to extract the customers' data which conform to particular customer profiles. For large data sets (in this case 200,000 instances), data mining methods supported by the ontology significantly simplified data analysis process. The developed solution can be easily customized to suit the analyst's needs - for example, when one needs to offer a new product, or to define a new class of customer profiles. To extract customer data for a specific marketing campaign, it is enough to enter the profile name (e.g., *High heels*) to get a list of persons who meet certain criteria.

An important direction for further research is therefore to improve the application interface by way of facilitating and monitoring the data mining process with special emphasis on the needs and skills of business analysts.

Acknowledgments. The authors would like to thank the staff of the VSoft company, Krakow, for providing the Pathfinder package, data, and documentations. The research was carried out as part of project No. POIG.01.04.00-12-106/12 - "Developing an innovative integrated platform for the financial area", referred to as the Project, co-financed by the European Regional Development Fund and the Innovative Economy Operational Programme 2007–2013. Special thanks to Ghislain Atemezing of Mondeca, Paris, for his helpful suggestions concerning the usage of Cellfie plugin and comments on draft of this paper.

References

1. Wyse, S.E.: Advantages and disadvantages of face-to-face data collection. Snap Surveys, 15 October 2014. http://www.snapsurveys.com/blog/advantages-disadvantages-facetoface-data-collection/. Accessed 25 July 2016

2. Pruitt, J., Adlin, T.: The Persona Lifecycle: Keeping People in Mind Throughout Product Design. Elsevier, Boston (2006)
3. Montgomery, D.B.: Marketing Information Systems: An Emerging View. Forgotten Books, London (2012)
4. Stair, R.M., Reynolds, G.: Principles of Information Systems. Cengage Learning, Boston (2016)
5. Nogueira, B.M., Santos, T.R.A., Zarate, L.E.: Comparison of classifiers efficiency on missing values recovering: application in a marketing database with massive missing data. In: Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining, CIDM, pp. 66–72 (2007)
6. Cao, L.B., Zhang, C.Q., Liu, J.: Ontology-based integration of business intelligence, web intelligence and agent systems. Int. J. **4**(3), 313–325 (2006). IOS Press
7. Matsatsinis, N.F., Siskos, Y.: Intelligent Support Systems for Marketing Decisions. Springer Science & Business Media, New York (2003)
8. Grassl, W.: The reality of brands: towards an ontology of marketing. Am. J. Econ. Sociol. **58** (2), 313–319 (1999)
9. Pinto, F., Alzira, M., Santos, M.F.: Ontology-supported database marketing. J. Database Market. Cust. Strategy Manage. **16**, 76–91 (2009)
10. Pinto, F., Alzira, M., Santos, M.F.: Ontology based data mining – a contribution to business intelligence. In: 10th WSEAS International Conference on Mathematics and Computers in Business and Economics (MCBE 2009), Czech Republic, 23–25 March (2009)
11. Zhou, X., Geller, J., Perl, Y., Halper, M.: An application intersection marketing ontology. In: Goldreich, O., Rosenberg, Arnold, L., Selman, Alan, L. (eds.). LNCS, vol. 3895, pp. 143–163Springer, Heidelberg (2006). doi:10.1007/11685654_6
12. Barbu, E.: An ontology-based system for the marketing information management (2006). http://clic.cimec.unitn.it/eduard/publications/OntologyBasedSystem.pdf
13. Saggion, H., Funk, A., Maynard, D., Bontcheva, K.: Ontology-based information extraction for business intelligence. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC/ISWC -2007. LNCS, vol. 4825, pp. 843–856. Springer, Heidelberg (2007). doi:10.1007/978-3-540-76298-0_61
14. Bouquet, P., Dona A., Serafini L., Zanobini S.: ConTeXtualized local ontology specification via CTXML. In: Bouquet, P. Harmelen, F., Giunchiglia, F., McGuinness, D., Warglien, M. (eds.) MeaN-02 AAAI Workshop on Meaning Negotiation, Edmonton, Alberta, Canada (2002)
15. Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R.V.: Semtag and seeker: boot-strapping the semantic web via automated semantic annotation. In: Proceedings of the Twelfth International WWW Conference (2003)
16. Domingue, J., Dzbor, M., Motta, E.: Magpie: supporting browsing and navigation on the semantic web. In: Nunes, N., Rich, C. (eds.) Proceedings of ACM Conference on Intelligent User Interfaces (IUI), pp. 191–197 (2004)
17. Linoff, G.S., Berry, M.J.A.: Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management. Wiley, New York (2011)
18. Ohsawa, Y., Yada, K.: Data Mining for Design and Marketing. Chapman and Hall/CRC, Boca Raton (2009)
19. Poh, H.L., Yao, J., Jasic, T.: Neural networks for the analysis and forecasting of advertising and promotion impact. Int. Syst. Account. Financ. Manag. **7**(4), 253–268 (1998). doi:10.1002/(SICI)1099-1174(199812)7:4<253:AID-ISAF150>3.0.CO;2-X
20. Witten, I.H.: Data Mining: Practical Machine Learning Tools and Techniques. Data Management Systems. Morgan Kaufmann, San Fransico (2011)

21. Tsiptsis, K., Chorianopoulos, A.: Data Mining Techniques in CRM: Inside Customer Segmentation. Wiley, New York (2009)
22. Pawełoszek, I., Korczak J.: From Data Exploration to Semantic Model of Customer, submitted to publication in ACM Transactions of Knowledge Discovery from Data (2016)
23. Vyncke, P.: Lifestyle segmentation: From attitudes, interests and opinions, to values, aesthetic styles, life visions and media preferences. Eur. J. Commun. **17**, 445–463 (2002)
24. Kahle, L.R., Chiagouris, L. (eds.): Values, Lifestyles, and Psychographics. Psychology Press, New York, London (2014)
25. Armstrong, J.S., Brodie, R.J.: Forecasting for Marketing. In: Hooley, G.J., Hussey, M.K. (eds.) Quantitative Methods in Marketing, pp. 92–119. International Thompson Business Press, London (1999). http://forecastingprinciples.com/files/pdf/Forecasting%20for%20Marketing.pdf. Accessed 17 Dec 2015
26. Chattopadhyay, M., Dan, P.K., Majumdar, S., Chakraborty, P.S.: Application of artificial neural network in market segmentation: a review on recent trends. Manag. Sci. Lett. **2**, 425–438 (2012). http://arxiv.org/ftp/arxiv/papers/1202/1202.2445.pdf. Accessed 17 Dec 2015
27. Yao, J., Teng, N., Poh, H.L.: Forecasting and analysis of marketing data using neural network. J. Inf. Sci. Eng. **14**(4), 523–545 (1998). http://www2.cs.uregina.ca/~jtyao/Papers/marketing_jisi.pdf. Accessed 17 Dec 2015
28. Prymon, M.: Marketingowe strategie wartości na rynkach globalnych. Wydawnictwo UE we Wrocławiu, Wrocław (2010)
29. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Data Management Systems, 3rd edn. The Morgan Kaufmann, San Fransico (2012)
30. Aggarwal, C.C., Reddy, C.K.: Data Clustering: Algorithms and Applications. Data Mining and Knowledge Discovery Series. Chapman & Hall/CRC, Boca Raton (2013)
31. Demsar, J., et al.: Orange: data mining toolbox in python. J. Mach. Learn. Res. **14**(Aug), 2349–2353 (2013)
32. Larose, D.T., Larose, C.D.: Data Mining and Predictive Analytics. Methods and Applications in Data Mining, 2nd edn. Wiley, New York (2015)