Customer segmentation utilization for differentiated approach

Jan Panuš, Hana Jonášová, Kateřina Kantorová, Martina Doležalová, Kateřina Horáčková
Faculty of economics and administration
University of Pardubice
Pardubice, Czech Republic
jan.panus@upce.cz

Abstract—**Customer segmentation and customer relationship management generally is interesting subject to explore. One of the main reasons why we should focus on exploring this type of research is to notion of customer loyalty and profitability to increasing market share. Customer value can provide basic information for more targeted and personalized marketing. In this paper, combination of ABC and RFM (Recency, Frequency, and Monetary) is used to customer segmentation. What is unique is using data mining segmentation method later on to dividing customers to four segments. We used well known Two Step, Kohononen – maps and K-means methods. Results show possibility to use these combinations for customer segmentation and results for different segments can be used to explain marketing strategies of company.**

Keywords—customer segmentation; data-mining; CRM

I. INTRODUCTION

Data minings methods can process the results of customer behavior analysis and can create base for manager access to customer relationship management (CRM). With the growing role of the customer, the importance of building the customer oriented companies is increasing [1].

Original transaction marketing is changing into marketing, based on building a relationship with the customer. The way of building the relationship with the customer has changed during the evolving history of the market. One of the customer approaches is the customer relationship management. Operating of customer relations is still going through changes, for example with the possible use of the social media [2]. For example, Kantorova, Bachmann are studying the level of the usage of data obtained through social sites for the building of CRM[2].

CRM is a managerial approach that helps with building and keeping customer relations on the mutually beneficial level. The target is to create permanent value for the company and also for the customer. It means the win-win strategy [4]. Long term profitability is always the main motivation for companies. In the CRM framework it means the integration of the internal processes and functions so that the obstacles are removed (lowering of the costs and barriers). Operating of all the processes is carried out with the customer in mind. The aim is the satisfaction of the customer by reacting to customer needs and by that, increasing customer loyalty [5].

The segmentation makes management's decisions easier. A company can decide which type of customers they want to target. It is mainly the selection of the groups of customers (segmentation of the market) that will be identified; approachable, sufficiently lagre, stable, perceptive, and will bring profit [6]. It is also important that the company can react to the selected segment based on its strong attributes [7]. That will help them to differ from other competitors. Generally the segment is defined as a group of consumers which is a combination of customers as similar as possible (homogeneous inside), and at the same time, boldly different to the characteristics of customers in other segments (heterogeneous outside), as described by [8, 9, 10].

II. PROBLEM FORMULATION

As [11] describe – For the effective running of the business, the evaluation of factors that affect the purchasing behaviour of customer is important. Knowledge of identical and non identical factors in the individual customer segments, allows companies to correctly determine the proper characteristics of the goods and choose effective marketing tools, mainly in advertising and marketing. That leads to the increase of the cost of these activities.

Companies have problems with situation how to deal with data they have about their customers. There are so many data in the companies and they have problem to be knowledgeable with them. Firms' don´t know what to do with them. There are some people who can join data with company and with public for firm's succes but there are just few. People who can combine data, analysis and technology are missing in most of firms. There are only 15% of Fortune top 500 firms who can analyse their data.

Currently more intensive use of the information technologies to collect customer date is visible. An interesting view of the market segmentation is presented by [12], when the satisfaction of customers within the retail banking service in Greece is being researched. It brings many findings, when the combining of the instrument elements for determining the satisfaction of customer and the dividing into the customer segments, is possible. The article confirms that customer satisfaction is the function of the quality of services, price, innovation and facilities to make the use of services easier. The aim of this article was also to find out

978-1-4673-8861-0/16/$31.00 ©2016 IEEE

if this is a model that exists between customer segments and customer preferences that require specific types of service. Reference [13] in their article brings new taxonomy of the differentiation between customers.

To manage the relationship between the customer and the organisation, it is possible to use many different methods from different fields. One of the fundamental domains is a data mining [14]. References [15, 16] present a review from the professional literature for the combination of different data mining methods for CRM. Given the fact that CRM often contains much data, data mining and its methods is a suitable tool for the executing of the analysis. An approach similar to the one in our article can be also found for example in [17, 18, 19] where the methodology of the market segmentation based on the demand and purchase is described.

Is it possible to solve CRM problem helped by clustering data mining methods? Is it possible to use for segmentation and for customer segments explaination combination of ABC, RFM analysis and clustering data mining techniques? We set up the main goal of this research on these research questions in such way: to group customers to segments with utilization of combination ABC, RFM analysis and clustering data mining techniques. Inovative benefit is connecting analytical methods used in marketing and in customer relationship management with data mining methods used for clustering. The reason why we made such combination is huge amount of data, for which data mining techniques are suitable. These methods can be used for customer grouping to segments by different attributes. These segments are base for aimed building of relationship with individual groups of customer within CRM.

In this article the customer segmentation (B2B) on realistic data is carried out. These were provided by an anonymous IT company offering advisory services in the field of internet marketing, creating of the web sites, e-shops, software for real estate companies, virtual tours etc. The company entered Czech market in 2004. The aim is to allow the selected company to effectively use the bonus program, which should achieve a more effective marketing campaign and increase of the satisfaction of their clients. To assist in achieving this aim, could be the dividing of the clients so the individual customer segments will show similar values in key attributes for a given type of business. The set up of different benefits will have an effect on each one of those segments.

Data mining tools from the CRISP-DM methodology in SPPS Modeler environment are used. Client segmentation is done based on the bonus of the client that is given by the combination of two CRM analyses – RFM analysis and ABC analysis. Selected analysis should evaluate the client according to the purchasing behaviour (RFM analysis) point of view, and from the other factors important for each given community type (ABC analysis).

III. PREPARATION OF DATA FOR MODELLING

The chosen IT Company has a modular information system. For our needs we used these modules:

Company 1 – contains the list of all clients. Attribute ID of the company is used for the analysis.

Company 2 – characterises clients according to the turnover. Attributes Monthly turnover (2013 to 2014) and when client paid “First payment“ were used.

Implementation – detailed information about all solutions. Attributes ID web and Activation date are used.

Orders – contains information about all orders. Every order fits in one of 15 offered projects. Attributes Type of the project Starting date, Profit, Cost are used.

Invoices – all issued invoices are recorded here. Attributes Amount, Date of issue and Date of payment are used for the analysis.

Before the actual processing the alterations are necessary because some attributes have blank values or the data recording is not accurate. These changes are carried out in MS Excel by using tools so the recordings can be worked on but the result will not be negatively affected. This phase of data cleaning is timely. Further the set of clients has to be adjusted because new customers show different behaviour than long-term customers. Only customers with history of payments older than 13 months are included.

The aim of the data mining is to find as little as possible of the compact clusters. That allows the execution of good quality interpretation for the bonus program. For the appraisal of the quality of findings Silhouettes measure is used. This combines principles of cohesion and separation of clusters. The data mining task is processed in the program SPSS Modeler Professional.

For the client analysis, it is necessary to extract from the current attributes, the new ones that will be characterised by:

- Profit according to the project – 15 different projects for the individual orders. For lucidity, the projects are separated in categories: main products (e-shop, www pages, software for Real Estate offices and e-learning) and others. To make sure that the length of cooperation will not affect this factor; only orders placed after 2012 are used. Output is Profit per client and project.
- Average time frame for the payment of invoices – for each client the Average is calculated. Time of the payment of the invoice considering the due date.
- Turnover trend – with the help of the linear trend function of the monthly turnovers, the new attribute Trend is created. This will place clients into categories (growth, stagnation around the value, stagnation around zero, decrease and lump sum payments)
- Number of provided upgrades of the web sites – attribute Number is connected to each client because one client can have more than one web site.
- Average time between upgrades – clients that haven't provided an upgrade yet are from the calculations eliminated due to the fact that this could have a negative affect. New attribute is calculated as an Average for the client and web site.

 978-1-4673-8861-0/16/$31.00 ©2016 IEEE

- Time since last upgrade – attribute Time is based on the date of the last upgrade.
- Length of the cooperation with the company - calculated from attribute 1_payment.

IV. MODELLING

After the previous adjustments 1567 IT Company clients that fit all defined restrictions are entered into next phase of the processing. Because the aim of the client segmentation is creating of the clusters according to the bonus value it is necessary to carry out the analysis to establish the value of the client.

A. ABC analysis

Attributes that characterise profitability according to the project, and are connected and maximized, were placed into groups A, B, C, D by the following calculations. To make clear how many percent of the overall sum value attribute is given. a new attribute Percentage was created. For the assessment of 20% of recordings (entries) that are 80% of the overall value of the total, it was necessary to create another attribute where the cumulative total is calculated. All entries with value of the cumulative total under 80% are placed in group A. Entries that are between 80-96% were placed in group B and those entries with cumulative total over 96% but under 100% are in group C. Entries where the cumulative entries equals 100% are in group D.

Fig. 1 shows the placement of clients into groups A, B, C, D according to the profitability criteria of main products. This finding exceeded the expectations of the Paret rule (20% of clients create 80% of profit). This is crucial information for the IT company.

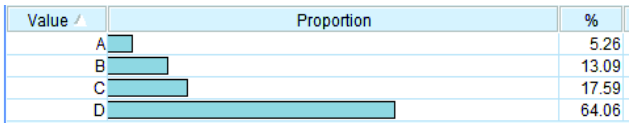

Figure 1. Customer segmentation by profitability of main products

For other attributes that have characteristics of the categorised data, and can't be placed into group A, B. C or D just by calculation, must be proposed by expert judgement.

Average time for payment of invoices in Fig. 2 shows category A with payment shorter than 7 days, category B with payments between 7 and 20 days, category C with payments to two months and category D where others customers are selected.

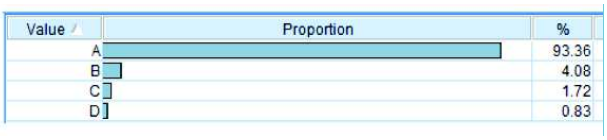

Figure 2. Customer segmentation by Average time of due period

The turnover trend is patterned on defined categories see Fig. 3. Group A means clients with growing turnover, group B means stagnation, group C means lump sum and group D means zero stagnation and descending trend.

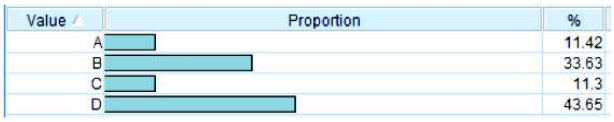

Figure 3. Customer segmentation by turnover trend

Fig. 4 shows groups by time of cooperation with company. Group A means clients with cooperation that is more then seven years, group B is six to seven years, group C is three to five years and group D are others customers.

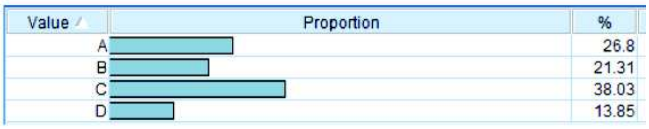

Figure 4. Customer segmentation by time of cooperation

We propose these rules for upgrade attributes:

- Number of upgraded products

If Number_upgraded_products_Sum >= 3 then
A
Elseif Number_upgraded_products_Sum = 2 then
B
Elseif Number_upgraded_products_Sum = 1 then
C
Else
D
Endif

- Average time between upgrade

If Number_upgraded_products_Sum = 0 then
D
Elseif Time_between_upgrade_Mean < 31 then
A
Elseif Time_between_upgrade_Mean < 50 then
B
Elseif Time_between_upgrade_Mean < 70 then
C
Else
D
Endif

- Time from last upgrade

If Number_upgraded_products_Sum = 0 then
D
Elseif Months_from_last_upgrade < 31 then
A
Elseif Months_from_last_upgrade < 50 then
B
Elseif Months_from_last_upgrade < 70 then
C
Else
D
Endif

Category A were evaluated by ten points, category B by five points, category C by two points and category D by zero point. The most important criterion for company is time from

last upgrade and the least important is number of upgrades from view of upgrading.

Maximum value of points per customer was set to 633.33 points. Customers with more than 50% of maximum value were ranked as group A, customers with more than 25% of maximum were ranked as group B, customers with more than zero-point were ranked as group C and others were ranked as group D – see fig. 5.

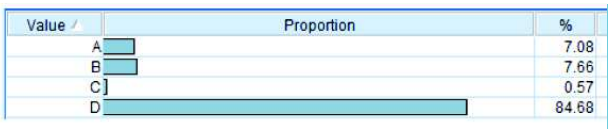

Figure 5. Customer segmentation by upgrade criterion

B. Criteria valuation

Each criterion offers client information that can characterise the behaviour and prospectively estimate future behaviour. For the actual segmentation and the overall valuation of the client, some information can be more important than others. Therefore it's important to evaluate each criterion and give them an importance weight that will affect the final result.

As the most important criteria are considered: criteria trend, profit according the project types. The values (importance, weight) were discussed with the management of the company, and we are convinced that these can be generally used for every IT company that has a similar orientation.

The following table No. 1 shows the individual categories and their importance. The entries will be placed into final groups A, B, C, D based on these. Category A is valued at 10 points, B 5 points, C 2 points and D was given 0 points.

TABLE I. VALUES OF CRITERIA USED FOR ABC ANALYSIS

Attribute	Weight	Attribute	Weight
Turnover trend	110	Average time of payment	70
Profit of main products	100	Profit training	60
Profit of internet marketing	80	Profit domain	50
Profit of webhosting	80	Profit technical support	50
Others profit	70	Length of cooperation	50
Average upgrade	70		

The calculation of the final evaluation of the client was done according to (1):

$$\frac{\sum_{i=1}^{n} v_i * k_i}{n} \quad (1)$$

where i is number of criterion, k_i is i criterion, v_i is weight for i criterion and n is number of criterion.

The minimal value that client could achieve is 709 points. The most successful customers received 692 points and the least successful had 303 points. Most companies received around 440 points.

C. RFM analysis

RFM analysis is based on the analysis of the individual entries of performed purchases. Analysis evaluates clients in three dimensions: recency, frequency and monetary.

The result of RFM analysis could be affected by the lump sum payments for the purchased services. Some customers pay lump sum monthly payments, some of them once every 6 months and some pay once a year. Clients with monthly payments could have an advantage so the lump sum payments are recorded as only one entry.

Firstly, the tool called RFM Agregate, calculates attributes of RFM Recency and RFM Frequency. Newly created attributes were valued in the bundle of the RFM Analysis according to importance. Most important is considered to be the spent amount value; second most important is considered the frequency of purchases. Least important is the time since the last purchase.

In the RFM analysis bundle, into how many segments in each individual attribute, will the recordings be divided, is also stated. Each criterion can have from 1 to 9 intervals. The original set up of 5 intervals was tested and the result showed that a better range is achieved by RFM Recency attributes when divided into 7 intervals (the frequency of purchases in IT company is not as high as other goods). RFM Monetary has 9 intervals (because of the large range of values) and RFM Frequency was left at the default setting of 5 intervals.

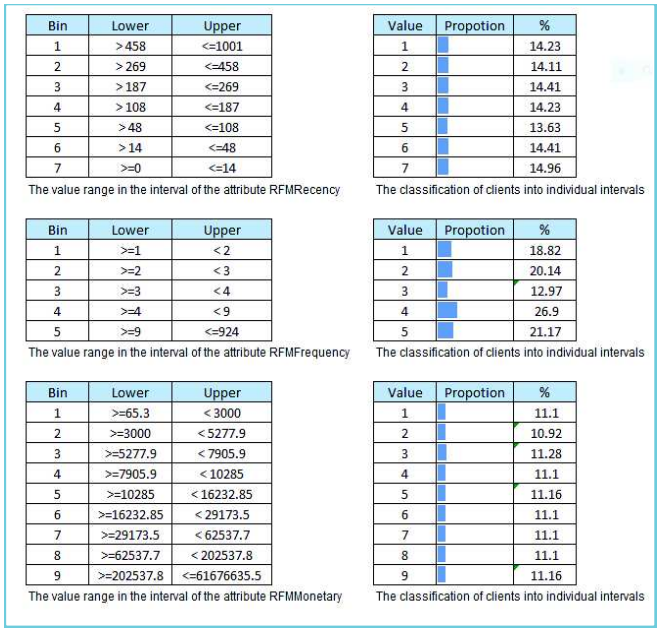

Figure 6. The value range in the interval of RFM attribute

978-1-4673-8861-0/16/$31.00 ©2016 IEEE

The result of given analysis is the RFM score that is stated by the RFM Analysis bundle and is calculated as:

$$RFM \quad L\ R \quad 4\ E\ R \quad (E\ R \quad /Æ \qquad (2)$$

where v_r, v_f and v_m are weights for variables recency, frequency and monetary. R, F and M mean number of points achieved per individual variables.

Maximum achievable score is 1040 points. 95 companies received the highest number of points (5,7% from the overall number of entries), most clients are in the middle - 595 points. Clients with 140 points had the lowest number.

D. Client segmentation

Known segmentation methods that are available in SPSS Modeler – Two Step method, Cohen maps and K-means – are used. The best results were achieved by the K-means method.

The elbow method was used to estimate the optimal number of clusters. The aim of this method is to select such a number of clusters K where quadratic deviation between the individual objects of given cluster, and its centroid suddenly decreases.

In SPPS Modeler, the model K-means, for 2 to 13 clusters, was created. The most obvious break is in 4 clusters. Silhouette value reaches the value of 0,51, which is a very good quality of the model. Created segments of customers are shown in Fig. 7.

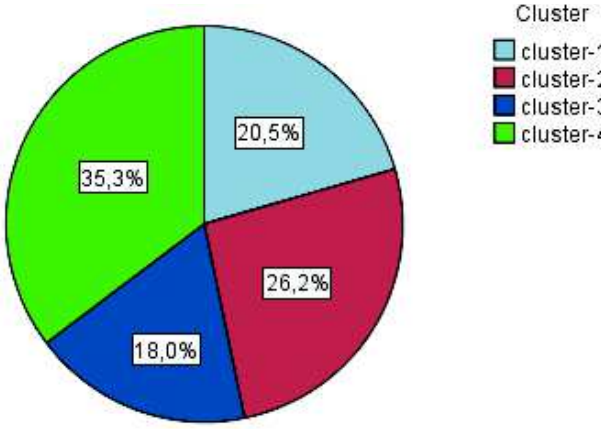

Figure 7. K-means methods and segments created

One of the aims is to make clear which segment is, from the point of client value, best and worst for the IT client. From Table 2, it is clearly evident that the best clients are in cluster 3, who achieved the average number of 707 points in RFM analysis and 497,7 points in ABC analysis.

Cluster	cluster-4	cluster-2	cluster-1	cluster-3
Size	35,3% (553)	26,2% (411)	20,5% (321)	18,0% (282)
Inputs	RFM Score 438,75	RFM Score 214,11	RFM Score 666,07	RFM Score 707,59
	Soucet 391,40	Soucet 371,51	Soucet 389,48	Soucet 497,77

Figure 8. Average of each attribute in the individual clusters

E. Characteristics of the created segments

Segmentation divided clients into relatively level homogenised groups that are different in the needs of the clients, their characteristics and purchasing history.

Best customers – cluster 3

18% of clients with the highest value for the IT company is in this group. The average spending of 1 065 657 CZK in given period has been achieved. This segment reaches the highest profit for the company in each project. On average the profitability of this segment is 5,6 times higher than in the second best segment.

They quite often upgrade their websites and at the same time, so as to make their products visible on the internet (project of the internet marketing). Customers from this segment are – in comparison to clients from other segments – very interested in multimedia and graphic services.

Good customers – cluster 1

Most of the clients in this cluster have only one product (in 95% it is a product without any type of integration). 86,6% of cases are clients with originally purchased products, who never upgraded their website. Almost 20% have none of their main product with the IT company and they use other services.

Average spending in the given period is 171 809 CZK. Average time for payment of invoices is 4 days after the due date.

Average customers – cluster 4

This segment has clients not quite with a zero spending, but a lot lower than other previous segments (14 471 CZK). These clients mostly purchased one of the main products and pay for the services of web hosting. They are not really interested in any other services. Clients in this segment have an excellent payment history.

Other customers – cluster 2

Clients that are not very lucrative for the IT company are in this segment. A typical client from this segment only spent 3 845 CZK in the last 2 years. There is no interest in investing money in their web sites or innovate them. 239 clients from this segment have purchased a main product (58,15%) but almost half of them are not „active“. It indicates that clients in this segment already use the services of the competitors, have financial difficulties and can't invest in their products.

F. Algorithm of client segmentation

After the segmentation, the IT company needs to be able to place new clients into one of the created segments. The below described algorithm should be able to help. The flow diagram graphically summarizes and describes the work that is necessary to do to place clients into appropriate segments. A client can be placed into one of 4 segments, that are created; or a client can be placed into the segment 'New customers'; or „Client can't be placed – missing data'. The problem may occur when there are no appropriate data for client´s classification. This fact is possible to eliminate within preparation of data so that we can have relevant data without possibility to shorten data file about

clients that we don´t know a thing. This paper does not address this issue but it would be interesting to focus on data verification. Fig. 9 shows flow diagram of client segmentation that we created for better calculation of combination CRM methods and data mining methods.

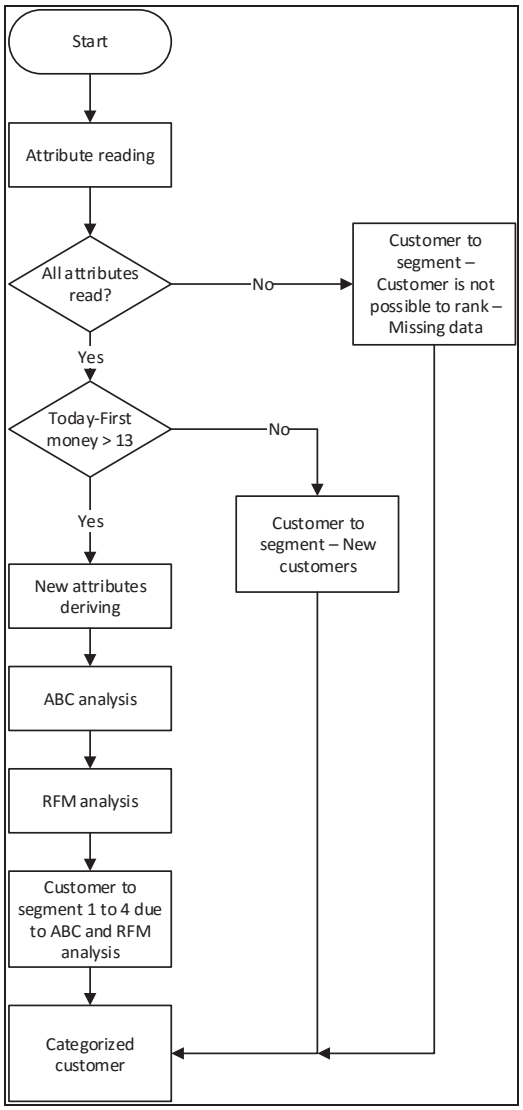

Figure 9. Flow diagram of client segmentation

V. CONLUSION

We determined these research questions for this paper: Is it possible to solve CRM problem helped by clustering data mining methods? Is it possible to use for segmentation and for customer segments explaination combination of ABC, RFM analysis and clustering data mining techniques? We set up the main goal of the paper based on these questions. Innovative contribution of this paper is to use combination of ABC and RFM analysis and data mining techniques for grouping customers to segments. Groups of customers were created for notion of customer loyalty and profitability to increasing market share.

We came out from information about size, frequency of purchasing, time of their last purchase, instalments of invoice etc. when dividing the customers to four main segments. First segment is called „Best customers" which is group that is active, with high rate of loyalty. We can recommend higher rate of investment to create relationship with customer in this segment. These customers bring high profit to the company and it is clear what type of products they are interested in.

Another segment is so called "Good customers". These customers usually bought only one product and they didn´t demand upgrade of this product yet. Company should propose some special offer of advantageous price for upgrade of product. It is possible to make proposition for main product of the company for 20% of these customers because they didn´t buy it yet.

Third segment is called "Average costumers". Strategy is similar as with segment of "Good customers". These customers are more likely conservative and they don´t have such good contribution for company. It is recommended not to invested many financial sources into this segment.

Last segment is called "Other customers". These customers don´t cooperate with the company almost. Their contribution for company is insignificant. Company shouln´t create any special campaign for these customers. It is possible to think about joining the campaign of another segment mentioned above.

Combination using data mining techniques for synthesis of data gained from ABC and RFM analysis is suitable for utilization within CRM approach to customers. These conclusions can be useful for companies with huge amount of data about their customers and they would like to use them by more efficient way of marketing and aiming on customer.

It is possible to consider with other data mining techniques such as decision trees or association rules for future research. It could be contribution for CRM to utilize this approach within the area of warranty, customer attraction or customer retention.

REFERENCES

[1] P. Setia, V Venkatesh and S. Joglekar, "Leveraging digital technologies: How information quality leads to localized capabilities and customer service performance," MIS Quarterly. 2013, vol. 37/2, 2015, pp. 565-594.

[2] A. Payene, Handbook of CRM: achieving excellence in customer management, Amsterdam: Elsevier Butterworth-Heinemann, 2006.

[3] K. Kantorová and P. Bachmanm. "From customer orientation to cocial CRM. New insights from Central Europe," in Scientific Papers of the University of Pardubice - Series D. vol. 36. Pardubice: University of Pardubice, 2016

[4] J. Veber. Management: basics, modern managers approaches, productivity and prosperity. Prague: Management Press, 2009.

[5] F Buttle. Customer relationship management: concepts and technologies. Amsterdam: Butterworth-Heinemann, 2009.

[6] M. Kotabe, and K. Helsen. Global marketing a management. Hoboken: John Wiley & Sons, 2008.

[7] P. Kotler. Modern marketing. Prague: Grada, 2007.

[8] P. Pelsmacker, M. Geuens, and J. Bergh. Marketing communication. Prague: Grada Publishing, 2003.

[9] V Bárta, L. Pátík, M. Postler. Retail marketing. Prague: Management Press, 2009

[10] J. Koudelka. Segmentation of consumer markets. Prague: Professional Publishing, 2005

 978-1-4673-8861-0/16/$31.00 ©2016 IEEE

[11] J. Novotný and P. Duspiva, "Factors affecting buying behaviour of consumers and their importance for enterprises," in E+M. Liberec, vol. 17, 2014, pp 152 – 166

[12] A. D. Athanassopoulos, "Customer satisfaction cues to support market segmentation and explain switching behavior," in Journal of Business Research. 2000, vol. 47, no. 3, pp. 191-207.

[13] T. Bock and M. Uncles, "A taxonomy of differences between consumers for market segmentation,". in International Journal of Research in Marketing. 2002, vol. 19, no. 3, pp. 215-224.

[14] V. Lendel and M. Varmus, "Proposal of Innovative Approaches of Relationship Marketing," in Business: Theory and Practice. Vol. 16, Iss 1, pp. 63-74, 2015.

[15] P. B. Henry. "Stock market liberalization, economic reform, and emerging market equity prices," in Journal of Finance. 2000, vol. 55, no. 2, pp. 529-564.

[16] E. W. T. Ngai, L. Xiu and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," in Expert Systems with Applications. 2009, vol. 36, no. 2, Part 2, pp. 2592-2602

[17] C. Y. Tsai and C. C. Chiu, "A purchase-based market segmentation methodology," in Expert Systems with Applications. 2004, vol. 27, no. 2, pp. 265-276.

[18] D. Bell, T. Ho and C. S. Tang, "Determining where to shop: Fixed and variable costs of shopping," in Journal of Marketing Research. 1998, vol. 35, no. 3, pp. 352-369.

[19] S. Simonova, „Identification Of Data Content Based On Measurement Of Quality Of Performance", In E+M Economics and Management, vol. 15, issue 1, pp.128-138, 2012.