

UDC 339.138

JEL Classification: M31, C45

<http://doi.org/10.21272/mmi.2019.1-20>

Atis Verdenhofs,

Riga Technical University, Latvia

Tatjana Tambovceva,

Dr. Sc., Professor, Riga Technical University, Latvia

## EVOLUTION OF CUSTOMER SEGMENTATION IN THE ERA OF BIG DATA

**Abstract.** *The development of information technology (IT) causes an increase in the amount of data to be created, stored and processed for the needs of various organizations. Segmentation is one of the marketing tools that can help the organization to promote sales activities and benefit from it. It is important for marketing practitioners and decision makers to understand the concept of predictive modelling and have an understanding of how to use big data for segmentation purposes. Marketing and Information Technology are blending due to digitalization, statistics is becoming more important due to the rise of big data and data mining opportunities. Borders of different disciplines are becoming vaguer and interconnection of disciplines can be observed more often. The purpose of the study is to create customer segments based on predictive modelling by using big data available in an organization. Data for modelling is used from a non-banking lending company based in Latvia AS 4finance. The process of data mining is described and performed in the study using data provided by the company. For the data mining process and the development of customer segments, the authors selected RapidMiner Studio software and used CRISP-DM data mining methodology. Three types of activities were tested to evaluate the economic benefit of created segmentation model on overall 11321 customers. All customers were segmented into two groups based on the created predictive model – one group contained customers that were predicted to become inactive and second group with customers that were not predicted to become inactive. All customers were split into three groups containing a similar split of predicted outcome. Three different types of activities were performed with all three groups. As a result, common characteristics of segmentation and predictive modelling were identified. The results of the empirical study show that it is possible to create customer segments by using sophisticated predictive model. This can be achieved without having to write statistical software codes. The study results also show that the organization can benefit from the implementation of segmentation based on data mining and predictive modelling in key business areas. Segmentation model created during research shows economic benefit for the company. Authors also indicate that this segmentation approach can be replicated in different business areas.*

**Keywords:** segmentation, Big Data, predictive modelling, decision tree, RapidMiner.

**Introduction.** Three disciplines are merged to create a new approach in decades-old marketing method – segmentation. It is important for marketing practitioners and decision makers to understand the concept of predictive modelling and have an understanding of how to use big data for segmentation purposes. In this research, the literature of segmentation, predictive modelling and big data are analysed, and all three dimensions are connected into one process that is described and performed on company data. The study is performed on data provided by non-bank lending company AS 4finance. The aim of the study is to create a predictive model that will segment customers into the ones that most likely will not become inactive and other customers. The outcome of segmentation is tested in sales activity and results described in this research paper. Results show that there is an economic benefit for the company to create such customer segmentation.

The aim of the research is to show interconnections between three disciplines based on literature from all three disciplines – marketing, statistics and Information Technology (IT). An actual segmentation model is created to support findings from the literature overview. The research question is following – can one create customer segmentation with the help of predictive modelling and big data that will generate an economic benefit to the company? The process of segmentation is performed with RapidMiner Studio software and is based on CRISP-DM methodology. Finding research is that widely used buzzwords like

---

**Cite as:** Verdenhofs, A., & Tambovceva, T. (2019). Evolution of Customer Segmentation in the Era of Big Data. *Marketing and Management of Innovations*, 1, 238-243.  
<http://doi.org/10.21272/mmi.2019.1-20>

«big data» and «predictive modelling» are decrypted and common characteristics with segmentation are identified. Interconnection is following – big data is the source of information for predictive modelling that generates outcome – one or several customer segments.

**Literature Review.** Segmentation started to become more widely accepted and used in the middle of the 20th century. Smith W.R. described in details segmentation and strategies that can be adapted by using segmentation. It is emphasized that «segmentation is essentially a merchandising strategy, merchandising being used here in its technical sense as representing the adjustment of market offerings to consumer or user requirements» (Smith, W.R., 1956). Quinn L. et al. (2007) indicates the main areas of development in market segmentation from the 1950s until 2000s. There are six main developments observed – one in each decade ending with the statement that customers became less predictable at the beginning of the 20th century. Another segmentation approach shows four different segments in the digital environment – business partner, skilled enthusiast, new experience seeker, close buddy (Kim, K.Y. and Lee, B.G., 2015). Goyat S. (2011) listed more than 15 benefits that companies can benefit from if segmentation is performed on customers.

The concept of segmentation in use is defined by Smith W.R. (1956), he states that «Segmentation is based upon developments on the demand side of the market and represents a rational and more precise adjustment of product and marketing effort to consumer or user requirements». Segmentation is the process when potential customers are divided into different segments or groups and this division is based on various criteria or characteristics. The analysis of different literature indicates that segmentation is a tool for marketing to promote sales activities, it has changed over time and companies can benefit from using segmentation. Shmueli G. defines what is predictive modelling and what is the predictive model. Predictive modelling is «the process of applying a statistical model or data mining algorithm to data for the purpose of predicting new or future observations». While the predictive model is defined as «any method that produces predictions, regardless of its underlying approach: Bayesian or frequentist, parametric or nonparametric, data mining algorithm or statistical model, etc.» (Shmueli, G. 2010). With predictive modelling one is able to predict outcomes by using statistical methods. It is often related to detection theory (Nevin J.A., 1969) because the theory has a similar approach with predictive modelling – have variables describing or determining the outcome.

Segmentation, as described by Smith W.R. (1956), is connected particularly with sales promotion and companies should adjust product offering for specific segments. At the same time predictive modelling definition does not include specific business processes or mentions business perspective at all – it is a process of creating predictions based on data. Predictive modelling is not bounded in any industry, sector or field of research. Therefore, it can be used also in segmentation that nowadays is related to the field of marketing. One limitation that authors of this research identify is that predictive modelling is still connected with the field of statistics and to perform segmentation based on predictive modelling some knowledge of statistics is required. Segmentation is concerned with selecting targeted customers, customer groups for specific offer or activity while predictive modelling produces new or future observations. This is the point of interconnection – new observations created by predictive modelling can be translated as customers with specific characteristics that will correspond to the defined target segment. For example, customers who will start using the product in a defined time period. Predictive modelling is an advanced form of creating customer segments.

France and Ghose (2019) investigated visualization, segmentation, and class prediction topics and emphasized links between these disciplines. Wong and Wei (2018) have done research on customer segmentation for online travel agencies. They segmented high-value customers, analysed their online purchasing behaviour and predicted their next purchases from an online air travel corporation. For these purposes, they developed a customer online behaviour analysis tool based on an integrated model of customer segmentation, customer behaviour and prediction analysis. The third dimension should be

added to marketing and statistics – Information Technology (IT) development with the rise of big data in particular. Big data is characterized by three «V» – volume, velocity and variety (Chen C.P. and Zhang C.Y., 2014). If these three characteristics can be found in data, then it can be defined as big data. Two factors that have intensified usage of big data among professionals and academics are access to powerful computers and user-friendly analytical software (Agarwal R. and Dhar V., 2014). Some authors indicate that in the case of predictive modelling there might be situations when the prediction is more important in comparison to an explanation of how the prediction is created. This, however, is not going to be the case in this research because the decision tree method is used for creating customer segmentation. Vincenzo Morabito (2015) in their book described the impact of Big Data and analytics on management and organizations, focusing on challenges for governance, evaluation, and change management. Verhoef, Kooge and Walk (2016) also underlined that Big data analytics is becoming a key driver of value creation in modern enterprises, wherein enterprise applications are designed to collect direct customer feedback and information from internal business operations. Big data analytics was examined also by Saggi and Jain (2018). They described various big data analytics tools, techniques and technologies. According to their research, there are seven V's characteristics of big data analytics: Volume, Velocity, Variety, Valence, Veracity, Variability, and Value. They divided the architecture of big data analytics for six components:

- 1) data generation;
- 2) data acquisition;
- 3) data storage;
- 4) advanced data analytics;
- 5) data visualization;
- 6) decision-making for value-creation and they can be used for various applications such as agriculture, healthcare, cybersecurity, and smart city.

Adrian et al. (2016) investigated the capability components for big data analytics implementation towards value discovery. Based on this investigation, they founded that «the capability components that may impact value discovery is formulating big data framework that includes the enabler technology and processing and using sufficient analytic techniques for analysing big data». Decision tree statistical classification technique is widely used because of the visualization option and ease of explanation to business representatives and decision makers. It is confirmed by an expert Howard J. in 2012 during a presentation in conference «O'Reilly Strata Conference. Making data work» when two commonly used types of predictive modelling «black box» and «explanatory» approaches are described. Similar findings of the easy explanation necessity are done by Muley P. and Joshi A. (2015) who researched customer segmentation techniques in a real-time environment.

**Methodology and research methods.** The software selected for data mining process is RapidMiner Studio that is freeware Graphical User Interface (GUI) type of software that does not require the need for writing statistical or another type of codes (like it should be done, for example, in another freeware statistical software widely used – R). There is more than one standard that can be followed when data mining process is performed (Azevedo A.I.R.L. and Santos M.F., 2008), however, the authors of this research used CRISP-DM methodology and followed 6 steps defined in this methodology.

1. Business understanding. Management of the company required to limit the number of customers who are offered discounts for product usage. Discounts are granted to customers based on very basic selection criteria while management challenges marketing that there are customers who would use product also without getting the discount. The target of customer segmentation defined – create customer segment with customers who will not apply for a repeated loan during 90 days after repayment of the previous loan.

2. Data understanding. Several groups of data are prepared based on available information. One group contains information that customers have provided, i.e., age, gender, email account domain,

language for communication, the bank used for transactions. The second group contains information about cooperation history of the customers with the company, i.e., the total number of product usage, days of being a customer, average days between product usage, additional product usage. The third group of information is external information received during the loan application assessment. It contains existence or non-existence of any negative credit history in any of credit bureaus used by the company. Each customer also contains a variable that splits customers into two groups according to the target segment according to their historical performance – they did or did not apply for a loan during 90 days after previous loan repayment.

3. The main concept to understand about predictive modelling process is that data used for predictions also must contain historical results of the target value. As in this case – prediction must show if customers will or will not apply for a loan during 90 days after repayment, then the prediction model is created on customers that historically have done target activities.

4. Data preparation. The authors of the research prepared dataset of 27 different variables used for modelling. Dataset consists of 13735 customers that are split into two parts for the next step in the process. From target variable perspective dataset is divided equally – half of the customers did apply for a loan during 90 days after the previous repayment, other half did not apply for a loan for 90 days. The first part is training dataset (6510 customers) – dataset on which predictive model is created and the second part is testing dataset (7225 customers) on which created a model is tested for accuracy.

5. Modelling and evaluation are performed in RapidMiner software since it requires an only limited amount of steps to be done. To create the predictive model three steps were done: the training set data retrieval; the setting of the target value (a variable indicating performance in 90-day period after loan repayment); adding a decision tree node for modelling.

To evaluate model additional steps are done: – testing set data retrieval; – model application node adding; – performance node adding. Total accuracy of the model is calculated in the performance step and show an overall accuracy of 77.56%.

6. Deployment phase is done after testing for the economic benefit that turned out to be positive. The company integrated segmented customer selection process into the business process.

**Results.** Three types of activities were tested to evaluate the economic benefit of created segmentation model on overall 11321 customers. All customers were segmented into two groups based on the created predictive model – one group contained customers that were predicted to become an inactive and second group with customers that were not predicted to become inactive. All customers were split into three groups containing a similar split of predicted outcome. Three different types of activities were performed with all three groups.

First customer group was not anyhow influenced by the company – this group did not receive an offer. Second customer group received a special offer in the form of discount for the next loan. All customers in the second group received offer despite predicted customer segment. Third customer group also received a special offer with an identical offer as for the second group with one difference – only customers that were predicted to become inactive received this offer.

Performance of the whole group was calculated to evaluate if the segmented approach shows economic benefit. The first indicator for comparison was the response rate of the group. As the response rate, in this case, researchers understand how many customers started to use the product during the 90-day period after the start of observation and/or issue of the special discount offer. First customer group that did not receive any offer showed a response rate of 1.41%, second group where all received special offer showed a response rate of 2.06% and third group where only predicted inactive customers received offer showed a response rate of 1.95%. Results show that by giving a special offer to all customers generates the highest response rate. To calculate economic benefit additional value is used, i.e. 12-month customer value. Response rates of all three groups are converted into a number of customers and

multiplied by 12-month customer value to compare income from customer groups over a longer period of time. Comparison of income from tested activity cannot be compared because the offer used for testing was 100% discount that generates no income at all for the second group. As a reference for comparison first groups result is taken – 100% as a base for income during the 12-month period. Second group show income of 105.6% compared with the first group while the third group show income of 115.6% compared with the first group. The results show that company can generate higher income by using segmented customer selection for discount offering and over a period of 12 months it will generate by 10% more income than if the special offer would be given to all customers. The second approach when the offer is given to all customers will generate higher product usage levels but will generate less income than a segmented approach. Any approach that company will select – either give discounts to all customers or give discounts only to selected customers will generate a higher response rate in comparison to no special offer approach.

**Conclusions.** Merge of three disciplines has created a new way of performing traditional customer segmentation. IT related phenomena big data can be used in predictive modelling (field of statistics) to created customer segmentation (field of marketing). Predictive modelling can be named as an advanced form of creating customer segments. To create segmentation based on predictive modelling a company must have historical data that corresponds to target outcome, i.e. to predict customers that will return to product usage a company must have such customers historically. Predictive modelling can be performed without the necessity of writing statistical software codes, by only using GUI type of software and traditional Microsoft Office software package (Microsoft Excel particularly). The whole process can be done by one person if he or she has an understanding of marketing offers, predictive modelling and has access to customer data and predictive modelling software. Coordination with decision makers and involved departments must be done during the process. This type of process can be replicated in many cases in the organization because the research was done only for one business process that was connected with giving a special offer to customers at one specific time period. More processes where customer selection is involved are identified and the process of predictive modelling and testing of outcome can be replicated if the company has the capacity to perform it. The authors assumed that discount offer to the customer was the main driver that impacted customer decision while other factors can be important as well, i.e. macroeconomic factors, different marketing messages used in the offering, additional activities performed for sales promotions. Such other factors were not evaluated and observed in this research. The authors indicate the area for further research – can another type of business process benefit from segmentation performed with predictive modelling. Sales activity can be directly measured by sales volumes and generated income while such area as customer service could be evaluated with other metrics.

**Author Contributions:** Both A.V. and T.T. developed a concept of the paper and the manuscript. T.T. supervised the project. A.V. derived the models and analysed the data. Both A.V. and T.T. contributed to the final version of the manuscript. T.T. worked out almost all of the technical details. A.V. and T.T. wrote the manuscript.

#### **References**

- Adrian, C., Sidi, F., Abdullah, R., Ishak, I., Affendey, L.S., Jabar, M.A. (2016). Big data analytics implementation for value discovery: A Systematic Literature Review. *Journal of Theoretical and Applied Information Technology*, 93 (2), pp. 385-393.
- Agarwal, R. and Dhar, V., 2014. Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research.
- Azevedo, A.I.R.L. and Santos, M.F., 2008. KDD, SEMMA and CRISP-DM: a Parallel Overview. *IADS-DM*.
- Chen, C.P. and Zhang, C.Y., 2014. Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data. *Information Sciences*, 275, pp.314-347.
- France, S.L., Ghose, S. (2019). *Marketing Analytics: Methods, Practice, Implementation, and Links to Other Fields*. *Expert Systems with Applications*, 119, pp. 456-475.
- Goyat, S., 2011. The Basis of Market Segmentation: a Critical Review of Literature. *European Journal of Business and Management*, 3(9), pp.45-54.

- Howard, J., 2012. From Predictive Modelling to Optimization: The Next Frontier. [Online] Available at: <https://www.youtube.com/watch?v=vYrWTDxoeGg> [Accessed 13 January 2018]
- Kim, K.Y. and Lee, B.G., 2015. Marketing Insights for Mobile Advertising and Consumer Segmentation in the Cloud Era: AQ-R Hybrid Methodology and Practices. *Technological Forecasting and Social Change*, 91, pp.78-92.
- Morabito, V. (2015). *Big data and analytics: Strategic and Organizational Impacts*. Springer. pp. 183.
- Muley, P. and Joshi, A., 2015. Application of Data Mining Techniques for Customer Segmentation in Real Time Business Intelligence. *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, 2(4), pp.106-109.
- Nevin, J.A., 1969. Signal Detection Theory and Operant Behavior: A Review of David M. Green and John A. Swets' Signal Detection Theory and Psychophysics. *Journal of the Experimental Analysis of Behavior*, 12(3), pp.475-480.
- Quinn, L., Hines, T. and Bennison, D., 2007. Making Sense of Market Segmentation: a Fashion Retailing Case. *European Journal of Marketing*, 41(5/6), pp.439-465.
- Saggi, M.K., Jain, S. (2018). A Survey Towards an Integration of Big Data Analytics to Big Insights for Value-Creation. *Information Processing and Management*, 54 (5), pp. 758-790.
- Shmueli, G., 2010. To Explain or to Predict? *Statistical Science*, 25(3), pp.289-310.
- Smith, W.R., 1956. Product Differentiation and Market Segmentation as Alternative Marketing Strategies. *Journal of marketing*, 21(1), pp. 3-8.
- Verhoef, P.C., Kooge, E., Walk, N. (2016). *Creating value with Big Data Analytics: Making Smarter Marketing Decisions*. Routledge: London. pp. 338.
- Wong, E., Wei, Y. 2018. Customer Online Shopping Experience Data Analytics: Integrated Customer Segmentation and Customised Services Prediction Model. *International Journal of Retail and Distribution Management*, 46(4), pp. 406-420.

**A. Верденхофс**, Ризький Технічний Університет (Латвія);

**T. Тамбовцева**, Dr. Sc., професор, Ризький Технічний Університет (Латвія).

**Еволюційний розвиток сегментації споживачів в епоху Big Data**

*Розвиток інформаційних технологій (ІТ) спричиняє збільшення масиву даних (Big Data), які необхідно генерувати, зберігати та обробляти для задоволення потреб споживачів. При цьому сегментація, як один із маркетингових інструментів, може сприяти розвитку торговельної діяльності організації та отриманню вигод від неї. Поряд із цим автори зазначають, що маркетологам та особам, які приймають рішення, важливо розуміти основні принципи прогнозного моделювання та аналізу Big Data з точки зору сегментації ринку. Результати дослідження свідчать, що маркетинг та ІТ поєднуються внаслідок діджиталізації, а статистика набуває все більшого значення завдяки зростанню Big Data та можливостей інтелектуального їх аналізу. Метою даного дослідження є формування принципів поділу споживачів на відповідні сегменти, використовуючи прогнозне моделювання на основі Big Data компанії. Вихідні дані для дослідження надані небанківською кредитною установою Латвії АТ «4finance». Таким чином, у ході дослідження, процес інтелектуального аналізу даних описаний та виконаний на основі інформації, наданої вищезазначеною компанією. З метою здійснення інтелектуального аналізу даних та виокремлення сегментів клієнтів автори використали програмне середовище RapidMiner та застосовували методологію інтелектуального аналізу даних CRISP-DM. У ході дослідження для оцінки економічного ефекту від створеної моделі сегментації споживачів (11321 споживача), автори перевірили три види діяльності компанії. У ході дослідження усіх клієнтів було поділено на дві групи відповідно до створеної прогнозно моделі: перша група складається із клієнтів, які згідно прогнозам, стануть неактивними, тоді як друга група представлена клієнтами, пасивність яких не передбачалась. Окрім цього, усіх клієнтів було поділено на три групи, які містять аналогічний розподіл за очікуваним результатом. Таким чином, на основі отриманих результатів дослідження автори виявили загальні характеристики сегментації та прогнозного моделювання. На основі результатів проведеного емпіричного дослідження авторами визначено, що створення споживчих сегментів є можливим за допомогою прогнозно моделі високої складності, не використовуючи при цьому статистичних кодів програмного забезпечення. Так, результати дослідження дають підстави стверджувати, що впровадження сегментації на основі інтелектуального аналізу даних та прогнозного моделювання можуть принести економічні вигоди організації у ключових сферах її діяльності. Таким чином, створена у ході дослідження модель сегментації демонструє економічні вигоди для компанії. Автори зазначають, що даний сегментаційний підхід можна використовувати у різних сферах бізнесу.*

Ключові слова: сегментація, Big Data, прогнозна модель, дерево рішень, програмне середовище RapidMiner.

*Manuscript received: 18.12.2018.*

*© The author(s) 2019. This article is published with open access at Sumy State University.*