universiteit hasselt

2014 | Faculteit Bedrijfseconomische Wetenschappen

DOCTORAATSPROEFSCHRIFT

# A comparison of the single-product offering approach to the multi-product approach. An application in retail banking.

Proefschrift voorgelegd tot het behalen van de graad van doctor in de bedrijfseconomische wetenschappen, toegepaste economische wetenschappen, te verdedigen door

**Katarzyna Andrzejuk**

Promotor: prof. dr. Koen Vanhoof
Copromotor: prof. dr. Benoît Depaire

D/2014/2451/14

## Quotations

*When I examine myself and my methods of thought, I come to the conclusion that the gift of fantasy has meant more to me than any talent for abstract, positive thinking.*

Albert Einstein

*Mathematics may be compared to a mill of exquisite workmanship, which grinds you stuff of any degree of fineness; but nevertheless, what you get out depends on what you put in; and as the grandest mill in the world will not extract wheat-flour prom peasecods, so pages of formula will not get a definite result out of loose data.*

Lord Kelvin, when speaking of the calculations of the age of the earth

*You do not concentrate on risks. You concentrate on results. No risk is too great to prevent the necessary job from getting done.*

Chuck Yeager

*We cannot teach people anything, we can only help them discover it within themselves.*

Galileo Galilei

# Acknowledgments

This thesis, which is the culmination of several years of work, would not have been possible to complete without the support and benevolent understanding from a broad group of people. Their presence, understanding and endless river of help allowed me to endure to the end despite the difficult moments, since creating and writing a doctoral dissertations is a great challenge, particularly when you have to combine scientific research with daily job responsibilities.

In the first place I wish to express my sincere gratitude to my promoter Prof. Dr. Koen Vanhoof and my copromoter Prof. Dr. Benoît Depaire for guiding me through the entire PhD project, which was one of the biggest projects in my modest life. You both introduced me to the Business Informatics area, helped to explore the countless twists and turns of this field of knowledge. Koen, I consider it an honor working under your supervision. You really helped me in keeping track of the overall thesis and gave me a real big credit of trust combined with extraordinary patience, forbearance, and immense understanding. I will never forget your positive criticism and kind advices. Benoît, you are the best supervisor I could ever imagine to have. You were always very helpful and available for fruitful discussions, feedback and really practical and useful explanations of unclear issues that helped me to overcome the problems within and beyond the scope of the thesis, which definitely clarified my doubts. You gave me the required motivators to develop and make myself mature to understand the research side of my work. Your professional attitude to work and responsibilities, your unusual thoroughness as well as your above-average meticulousness and inestimable scrupulousness will always aroused my respect. Without you and your relevant and objective criticism this research would not have been completed.

At this point, I would like to gratefully acknowledge my collegues from the Bank, Mirela Starosielec, Karolina Bąk and Wojciech Werochowski for giving me the incentive to start writing and for believing that I could succeed. My special gratitude goes to Mirela for her continuous support and endless faith in bringing this project to a successful end.

Furthermore, even if I cannot find the right words - I want to express my appreciation for the support of my family – Mum, Dad, Sister, Auntie and my Son. Thank you for your unwavering faith in me, your boundless love, for the many conversations that gave me new strength to work and, in particular, for all the sacrifices that you had to bear because of me and the fact that you believed in me the whole time. You let me develop myself as a person and gave me the strength to cope with the big and small adversities. For all those mentioned and those unspoken great and little things I want to thank you more than I can express.

I cannot forget my friends, who I had a pleasure to meet in Diepenbeek. Without you, Fortunato Piersimoni, Emmanuele Del Fava, Carolina Medina Gomez, Donato Spoltore and Justyna Bakowska my time spent in Diepenbeek would not have been so happy and full of laughter, nice memories and Italian tastes. And Belgium would not have been so sunny, even during those rainy days. I am very glad that the Grand Design has made you a part of my PhD life. Thank you.

And last but not least, I would like to thank my professor from Warsaw School of Economics, Dr. Jerzy Surma for inspiration and for providing the constructive remarks.

Furthermore, I would like to thank all members of the Jury for their thoughtful and valuable remarks. It is a true privilege to have such respected members of the academic community as members of my PhD Jury.

*.Katarzyna Andrzejuk, 2013*

# Table of contents

# List of Figures

# List of Tables

# A multi-product offering approach as the first step to change product-centric company into the customer-centric. An application in retail banking.[1]

[1] This is the new title of the thesis, which was agreed after the whole thesis had been completed. In the opinion of the author this title is more adequate to the text of the thesis.

# 1.Introduction

## 1.1 Background and the company challenges

**A customer**, or more specificly, a consumer (from the Latin word "consumens") in the common sense is the person who consumes or buys goods and services, or otherwise – he/she is a link that appears at the end of the economic chain. But the most important attribute of the consumer is that in addition to the manufacturer or company, he/she is the main player in the market. The relationship between these two entities (consumer and company) is the essence of the market. In economics and marketing, statistical inferences on a consumers are most frequently carried out at the level of groups of consumers rather than on the individual consumer, especially when the focus is laid on investigating is a statistical concept, which is not analyzed individually but as a set of consumers, in order to know the trends and typical behavior. Moreover, it is very common that the customer, who has been once acquired by a company, is regularly targeted by this company with different marketing offers to let him/her become more loyal and to **expand his/her product portfolio**. This action, which is called **cross-selling** is a fundamental element for the customer lifetime cycle and it is able to produce a greater effect on the total than the sum of its parts (Rust and Chung, 2006). Although an investigation of prior literature related to the term "cross-selling" clearly reveals that there is no common definition of this concept (Shäfer, 2002), the simplest and widely known definition implies that cross-selling is the practice of sellnig an additional product or service to the existing customer. On the other hand, it needs to be mentioned, that depending on the industry, buyer or seller perspective, relationship status, time of purchase and organizational alignment, “cross-selling” can mean many different things. But for all of these kinds of organizations cross–selling enables the organization to increase its sale volumes and to exploit the company`s product portfolio with no extra costs in production or distribution. As a result, in general, cross-selling allows the company to generate higher margins while using less capital per dollar of sales and enables longer and deeper customer relationships (Harding et al., 2004). Among other reasons, it is easy to observe that nowadays cross-selling ranks as a top strategic priority for many industries including financial services, insurance, health care, accounting, telecommunications, airlines and finally – retailing.

By selling additional products and services from the company`s portfolio to existing customers, the cross-selling transaction creates more value not only for the company but also for customers. When the tailored offer meets the customer`s expectations, the customer will reduce the number of firms from which he/she buys, which simplifies the buying process, makes the customer satisfied and may encourage stronger cooperation with a specific firm (Homburg, Kuester, 2001). A satisfied customer becomes a sustainable competitive advantage of the organization (Fader et al., 2006; Kumar et al., 2009; Shah et al., 2006).

Despite all the benefits of cross-selling, many companies still encounter economic and financial challenges in trying to reach their cross-selling potential (Mundt et al., 2006). Firms usually focus on core processes rather than on customer needs, which generates a great potential for conflict (Belz, 1999). This phenomenon has been confirmed by **Galbraith** (2005), who is a leading expert on global organization design, in his research. In most cases he investigated, the sales force has been structured according to the product areas and individual business units. People responsible for sales have had clear product obligations and incentive schemes that have focused their efforts on one product area of business units. Such organizations were diagnosed as being **product-centric**, although their owners were convinced that their company is **customer-centric** because they owners had been working for years to understand and please their customers (Galbraith J., 2005). At the same time, in most industries, it is difficult to make money just by selling single products irrespective of the size of the portfolio the firm possesses. Stand-alone products and services commoditize rapidly and reduce profit margins. Looking for the remedy for such occurrence some scientists (Seybold, 2001; Selden and Colvin, 2003) suggest that companies will be evaluated on the basis of the total value of their customer relationships, which has been supported by results from studies comfirming the fact that sales to existing customers are more profitable than sales to new customers. Additionally, from a theoretical point of view, customers and their relationships with a company have been considered as valuable organization **assets** for decades (Smith et al., 2006; Berger et al., 2002; Blattberg et al., 2001; Gupta and Lehmann et al., 2003). Moreover, it becomes a more popularbelief that firms need to organize around loyal customers, whose relationships with the company should be properly managed in order to gain the highest company effectiveness. Different companies want to do business differently and being a profitable company means having the capabilities that allow for malleability. It also means forming long-term relationships with the most valuable customers, and interacting with valuable customers across multiple points of contact and integrating the results of these contacts into a consistent company position. Then, it means learning from the history and customer contacts to customize the company`s offering for different customers and learning about customer needs and expanding the company`s offering to meet them. It finally means using gained knowledge of customers to prepare products and services into solutions that are needed by customers and that create proper value for given customers. All of these aspects lead to the impression that to realize cross-selling a company needs to transform from a product- to a customer-centric organization. By avoiding the commoditization of products, customer-centric companies are able to offer solutions to customers, instead of stand-alone products. In this evolution mass customization and customer involvement trends have emerged (Piller, 2011), because firms aim to provide product bundles and integrated different products from different business units (Davies et al., 2000; Foote et al., 2001; Gulati, 2007; Tuli et al., 2007). Moreover, customer-oriented companies are interested in maximizing customer profitability and use customer relationship management as a most important process in order to manage the customer portfolio. In general, companies that use their resources to establish a customer orientation leverage their resources more effectively than product-oriented companies (Coviello et al., 2002; Day, 2006; Gulati and Oldroyd, 2005; Kohli and Jaworski, 1990). Table 1.1 presents all the differences between the product- and customer-centric organizational forms.

*Table 1.1 Product-centric organization and customer-centric company according to the Galbraith (2005).*

| | **Driver** | **Product-centric organization** | **Customer-centric organization** |
|---|---|---|---|
| Strategy | Goal | Best product for customer | Best solution for customer |
| | Main offering | Specific products | Personalized packages of service support, education, consulting |
| | Value creation | Cutting-edge products, useful features, new applications | Customizing for best total solutions |
| | Most important customer | Most advanced customer | Most profitable, loyal customer |
| | Priority-setting basis | Portfolio of products | Portfolio of customers-customer profitability |

| | Pricing | Price to market | Price for value, risk |
|---|---|---|---|
| Structure | Organizational concept | Product profit center, profit reviews, product teams | Customer segments, customer teams, customer profit and loss statements |
| Processes | Most important process | New product development | Customer relationship management |
| Rewards | Measures | Number of new products, Percentage of revenue from product less than two years old, market share | Customer share of most valuable customers, customer satisfaction, lifetime value of a customer, customer retention |
| People | Approach to personnel | Power to people who develop products: highest reward is working on next most challenging product, manage creative people through challenges with deadline | Power to people with in-depth knowledge of customers` business, highest rewards to relationship managers who save the customer`s business |
| | Mental process | Divergent thinking: How many possible uses of this product? | Convergent thinking: What combination of products is best for this customer? |
| | Rewards | Based on business unit performance | Based on company performance |
| | Sales bias | On the side of the seller in a transaction | On the side of the buyer in a transaction |
| | Culture | New product culture: open to new ideas, experimentation | Relationship management culture: searching for more customer needs to satisfy |

From a practical point of view, the managers are looking for a way to change the company from product-centric into customer-centric in order to make their company competitive and try to win the market. In order to become customer-centric organization, managers very often have to undergo a strict change process which often confronts them with risks and uncertainties associated with establishing cooperation among business areas and profit centers. Creating the customer-facing organizational units is a big challenge as it needs to change not only the structure, but also the management belief that the company is already customer-centric and does not need any changes. The biggest contrast between the product-centric and customer-centric organizations, which are the easiest to observe from business practices, is that a product-centric company tries to find as many **uses and customers as possible for its product**. So, from a marketing point of view this approach is the same as creating a **single product offer** which is proposed to customers in a marketing campaign. As far as a customer-centric company is concerned, this kind of company tries to find as many **products as possible for its customer** and it has to integrate those products. Thus, translating it into a marketing approach seems to be the equivalent of constructing a **multi-product offer**, which is proposed to the customer in the marketing campaigns. Changing the type of customer offering seems to be a first step into changing the company’s organizational form.

## 1.2 Purpose, research context and corresponding research questions

The **purpose** of this study is to explore how to **start changing** the company organization from a product-centric to a customer-centric company by changing the marketing offering from a single-product offer into a multi-product offer. Thus, unlike to the typical statistical doctoral thesis, the focus does not lie on the selection and optimization of data mining methods used to support the cross-selling process of a given company or to calculation of the cross-selling campaign potential. The focus rather lies on presenting the way of using the well known statistical and data mining methods such as logistic regression or K-means clustering in order to find the solution of a multi-product offering campaign construction and secondly, on presenting the **real case study** design and its results. It needs to be mentioned that the real case study often coincides some **consequences**. A first consequence encompasses the standard company procedures that had to be followed and the data mining methods that were used. The second consequence results in the neccessity to involve managers in the process of constructing the framework of the multi-product offering campaign. The last consequence provides the possibility to build the real life experiments on the market in the form of a champion and challengers group of customers. The market in this context means the financial market, and more specifically, one of the retail banks in the East-Central Europe. Although the first two consequences also constitute simultaneously the **limitations** of the research, the third one represents the **opportunity** and considerable advantage compared with a typically scientific thesis.

For most companies cross-selling is an option, not a necessity (Fleming, 2006; Gulati, 2007). Therefore, when a company decides to implement a cross-selling strategy, the managers invariably face high risks and potential costs due to uncertainty and changing market conditions, here financial market conditions (Duclos et al., 2008). By looking for the methods to optimize costs and profits, research can identify ther implications and ways to reduce costs and risks while reaching the same high profits. The most common challenge of cross-selling is to know which product to target to which customer. Typically, a company has several candidate products but, unfortunately, according to most managers, it is impossible to target all of these products to each customer. According to them, this may be too expensive, too time consuming or ineffective due to the information overload on the customer. Or, the company may not want to turn off the customer by flooding him/her with too many offers. Moreover, in the financial companies managers want to meet the customer needs in the best way and they want to be very precise in offering the customers with the right product they need at a given moment. In order to gain this precision in addressing customers with the right products, they often use data mining methods. These methods help to describe customers and make it easier to to predict the probability of a buy given a bank product for each customer. The most frequently adopted method which could be met in the financial institutions using cross-selling strategy supported by CRM systems and data mining methods is a next-product-to-buy (NPTB) model which promises to enhance the effectiveness of cross-selling by specifying which product to target to which customer. The NPTB model predicts which product a customer would be most likely willing to buy, given what could be known about the customer. To optimize the product portfolio it is necessary to build the NPTB model for each bank product and according to all the border conditions (e.g. company budget, staff capacity, company strategy, product priority selling and volume plans) in order to find a way to prepare a single marketing offer for each customer. On the other hand, also multi-category models exist, which are of interest not only to retailers but also to packaged goods manufacturers (e.g. Procter and Gamble), who sell products in multiple product categories (such as detergents, shampoos, etc.). However, such models are also of interest to firms such as financial services providers, who are interested in undertaking cross-selling initiatives across product categories (Kamakura et al., 1991). Recall that the most important customer-centric approach assumes multi-product campaigns instead of the single product offering. Thus, this study highlights the way to change, design and realize multi-product offerings successfully. In short, the main research question this study aims to answer is: **How to change the product-centric into the customer-centric by changing single-product offer into multi-product offer?** In order to answer this research question, this dissertation first presents a theoretical background of the business issues and literature review of Customer Relationship Management as a basic entertainment where marketing campaigns are prevalent, of the cross-selling strategies responsible for creating marketing offers in particular and methodological discourse of basic data mining methods. The questions answered are: What does the company know about the customer? What is the main definition and main idea behind the CRM strategy as a key process in a customer-centric organization? What possible type of CRM is available to meet in the company? What does the term "valuable customer" mean? What type of benefit is possible to create by cross-selling strategy? What does the Knowledge Discovery and Data Mining framework look like? In the next chapter a newly proposed methodology for a multi-product offering is presented, answering in detail the question how to find a solution to prepare a multi-product offer which suits each customer, based on the NTPB models (presented in Chapter 3). Afterwards the question – as to how to construct a multi-product offering campaign will be raised? The answer to this question will be given in the Chapter 4, which focuses on the multi-product offering campaign design. The last research question, which will be answered in Chapter 5, concerns the way to evaluate a multi-product offer campaign. It will also be determined which approach gives better results.

In conclusion, in order to provide an answer to the main research question, this dissertation follows a cumulative research approach with logical multi-product offering introducing process starting from the idea and theory of building the multi-product campaign, the way of construction and design, application, and ending with results from the real experiment.

In summary, all main sub-questions are listed below:

1. What is the reason of changing the company organization from product-centric into customer-centric orientation?
2. What is the main definition and conceptual framework of CRM from a theoretical point of view?
3. What role does cross-selling strategy have in a customer-centric company?
4. What does the company know about the customer and what is the theoretical framework for understanding the Knowledge Discovery and Data Mining (KDDM) side of the research.
5. How to successfully build the most optimal model to be used in the multi-product offering company?
6. How to develop and construct a multi-product offering campaign?

7. How to evaluate a multi-product offering campaign?
8. Which approach (single or multi-product) offering gives better results?
9. What are the conclusions and recommendations for further research?

From a methodological point of view, looking for the answers to the questions mentioned above, the objective of this research is to present a new combination of econometrics, widely known methods, to define groups of customers who possess similar likelihood to buy bank products.

The research title underlines that this study presents a new marketing offering approach, and verifies whether this new approach (a new type of multi-product offering) performs better than the common approach (a single-product offering) in a predefined context. Secondly, the focus lies on changing the approach, not by changing everything and starting everything from scratch, but by using all available information in the company and optimizing the analyses to provide a new picture of the customer, focusing on customers rather than on products. The third delineation of this dissertation is the analytical part, which explains the way to change the approaches. Achieved in a real business case study, the results are compared and complemented with remarks and recommendations for further deployment, improvement and application in the business strategy.

## 1.3 Outline of the dissertation

In the previous section, the research questions and motivation were clarified. In this section, the general structure of the dissertation is given.

The dissertation is divided into **two main research parts**, preluded by an introduction which provides a descripton of the research problem, the structure and the overall contribution of the thesis and recapitulated by a final chapter as it is shown the **Figure 1.1**. In the first part, the business and methodology background is examined. Chapter 2 deals with business issues which concern environment where the new approach is tested. It also gives a literature overview of the issues described. Chapter 3 presents the methodology and analytical framework according to which all the analytical tools were built (KDD process and logistic regression modeling). In the second part of the study, there is a need to test the proposed hypothesis from the first part of the dissertation, introducing the multi-product offering campaign. This is realized by finding similar groups of customers with their characteristics of buying needs. Thus, Chapter 4 covers the real case study and application of the proposed approach. It describes all available data and created propensity-to-buy models. It also shows the segmentation details and how the multi-product offering test has been constructed with all maintained assumptions. In Chapter 5 the results from the experiment and comparisons between single-product and multi-product offer are described. To conclude the thesis, Chapter 6 recapitulates the most important findings of the project along with overarching conclusions, theoretical and managerial implications and limitations and recommendations for further research.

*Figure 1.1 Overview of the structure of the dissertation.*

***1. INTRODUCTION***

**RQ1**: What is the reason of changing the company organization from a product-centric into customer-centric orientation?

***PART 1***

2. BUSINESS ISSUES AND THEORETICAL FRAMEWORK

**RQ2**: What is the main definition and conceptual framework of CRM from a theoretical point of view?

**RQ3:** What role does a cross-selling strategy have in a customer-centric company?

**RQ4:** What does the company know about the customer and what is the theoretical framework for understanding the Knowledge Discovery and Data Mining (KDDM) side of the research?

3. A METHODOLOGYYFOR MULTI-PRODUCT OFFER MARKETING CAMPAIGN

**RQ5**: How to successfully build the most optimal model to be used in the multi-product offering campaign?

***PART 2***

4. CASE STUDY - DEVELOPING THE MULTI-PRODUCT OFFERING CAMPAIGN

**RQ6:** How to develop and construct a multi-product offering campaign?

5. CASE STUDY - EXPERIMENTAL EVALUATION

**RQ7:** How to evaluate a multi-product offering campaign?

**RQ8:** Which approach (single or multi-product) offering gives better results?

***6. CONCLUSIONS***

**RQ9:** What are the conclusions and recommendations for further research?

# 2. Business issues and theoretical framework

The financial business sector has changed over time from being transaction centric to relationship-centric (Kumar et al., 2009). **Retail banking** is a part of the financial business, so these changes are also present there. Business specialists now see the merit in nurturing and growing profitable long-term relationships with their customers. Customer Relationship Management (CRM) has grown considerably in recent years, specifically to deal with this challenge (Jain and Singh, 2002). Therefore, this chapter contains the description of two frameworks: Customer Relationship Management and Data Mining methodology from two corresponding perspectives – the business perspective and the analytical perspective.

## 2.1 Business framework

### 2.1.1 Customer Relationship Management

Given the close connection between the current Chapter and Chapter 1 it is worth quoting at this point one of the golden thoughts of Gandhi: *"A customer is the most important visitor on our premises. He is not depending on us. We are depending on him. He is not an interruption of our work but the purpose of it.He is not an outsider but a part of it. We are not doing him a favor by serving him. He is doing us a favor by giving us the opportunity* *to do so."* So treating this sentence as a key definition of the most important entity for all customer centric strategies, it is easier to enter into the CRM world.

According to Dr. V. Kumar (2010) the process of customer management has a long history, and not only in the field of financial business. Its development has evolved through personal contacts with customers, towards an overview at the hole market, followed by customer segmentation and finally, however paradoxically, a return to the original approach – a personalized contact with the customer by using **interactive marketing** and analytical systems. In the interactive approach the customer is in the center, everything happens around him/her and everything depends on him/her. He/she is treated as an independent entity with specific needs, in a particular place and time (Deighton et al., 1996). On the other hand, to win in a highly competitive business environment it is necessary to a company and its managers to be able to identify profitable customers and to build long-term relationships with them (Jain and Singh, 2002). That is also a reason for the fact that marketing strategies which are based on the philosophy of **relationship marketing** and the related CRM (customer relationship management) methodology have gained a huge popularity (Palmatier et al., 2008) in recent years. When tracing this strategy in the banking sector, it has become evident for almost each company that from a marketing point of view it is not only important to acquire new customers, but equally important to build lasting relationships with current customers and maximizing the benefits of certain relationships as well. By building a personal, lasting contact with the customer, the bank can try to increase the customer value. Although the term "customer value" is usually different for every company and the importance of its ingredients may vary across companies, the immutable components of this concept for financial enterprises are customer loyalty and profitability. From a literary point of view a **valuable** customer is very often treated as a firm **asset** (Anderson et al., 1994; Gupta er al., 2006; Levitt, 1983; Smith, 2006), which means that for every type of company this is a person who brings extra income and whose relationship with the company is **long-term** (Morgan and Hunt, 1994). But from the practical perspective, it was found that the occurrence of last meaningful changes in a business environment had led to appear a highly competitive financial market. Since the number of the competitors is great and each of them tempt customers with wide product offer it caused the situation, that the customers are less loyal. Easy access, a high level of customer awareness and a very range of bank product portfolio made the fight of the customer increasingly difficult from companies' perspectives. The business strategy, which was widely used until recently, often boils down to minimizing costs and maximizing profits. This strategy, by the company being regarded as a customer-centric approach, very often occurs to be product-centric. Such solution is not sufficient in any condition to obtain and maintain a competitive advantage. The company has to compete for customers by focusing attention on those ones who can create the most benefit. Therefore, the identification of customers who can bring the greatest benefit for the company is very important. In addition to the profit which the company can obtain by selling to a particular customer, customers` loyalty and willingness to recommend the institution to customers` friends, meaning potential new customers may prove to be at least equally important, or even more important. The easiest way to identify the most valuable customers is to calculate the current value of sales for a particular customer and to use this to forecast future earnings. However, this indicator may incorrectly indicate the overall value of the customer. A customer who buys only a small number of products might recommend the company to large number of his/her friends. Conversely, a customer who buys a lot of company products noay not recommend a company to others at all. Therefore, regardless of sales volume, all customers should be treated as well as possible, so that they are satisfied with the contact with a given institution. This approach requires the treatment of each client separately, and individual responses to their needs.

From a theoretical point of view, especially those who have accepted the role which customers play in the creation of the value of the company, strange sentence structure customers and their relationships with the firm should be treated as critical resources that contribute to a competitive advantage for the company (Srivastawa et al., 1998). That is the main idea behind the **market-based assets** perspective, which assumes that the customer is a one of company resources, and who is valuable, rare, inimitable and non-substitutable (Barney, 1991), who needs to be utilized to get superior performance (Collis and Montgomery, 1995; Grant, 1991; Wernerfelt, 1984) and should also be developed, augmented, leveraged and valued in a similar way to the traditional resources (Srivastawa et al., 1998).

**Customer Relationship Management** (CRM) is the implementation of relational marketing, which, unlike the traditional approach, puts the focus on building lasting customer relationships (Morgan and Hunt, 1994), which is regarded as the basis of profit in the long term, rather than maximizing short-term sales (Reichheld and Sasser, 1990) and is presented as a main process in the customer-centric approach (Gilbraith, 2005). In the literature CRM is treated as the management of a mutually beneficial relationship from the perspective of the seller (LaPlaca, 2004), who benefits from all those in the relationship (Mitussis et al., 2006). Value is created for both, the customer and the company during cooperative and collaborative relationships. According to Swift (2001), CRM is the enterprise approach aimed at understanding and influencing customer behavior in order to improve customer acquisition, retention, loyalty and ending with customer profitability. For a more specific **CRM definition**, Payne and Frow (2005) could be quoted – CRM is *"a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. CRM unites the potential of relationship marketing strategies and information technology to create profitable, long-term relationships with customers and other key stakeholders. CRM provides enhanced opportunities to use data* *and information to both understand customers and co-create value with them. This requires a cross-functional integration of processes, people, operations, and marketing capabilities that is enables through information, technology, and application"*. In order to quote a more business oriented definitions of CRM the one which was developed by Hobby (1999) is the most appropriate: *"CRM is a management approach that enables organizations to identify, attract, and increase retention of profitable customers by managing relationships with them"*.

Even though awareness of customer importance in the trade world was known from the very beginning, the business strategy based on CRM started to appear around twenty-five years ago. It was Dwyer et al. (1987) who stressed the role of the relationship aspect of buyer-seller behavior. Afterwards some other scientists (Reichheld, Sasser, 1990) proved that the companies focusing on relationships may obtain significant advantages because customers tend to generate higher profits with higher company loyalty. Finally, most common definitions were classified by Richard and Jones (2008) into **two related categories**:

1. CRM is often defined as a form of **relationship** strategy, e.g. *"CRM is a comprehensive strategy and process of acquiring, retaining and partnering with selective customers to create superior value for the company and the customer"* (Parvatiyar and Sheth, 2001). When applying it to the research context of this study, this definition describes the structure of a CRM team, which takes care of the customer from every possible prespective. There are managers responsible for cross-selling, for retaining and everything is connected by the link to the analytical profile of the customer, who if he/she is offered in a proper way and creates superior value.
2. CRM is also often described from a more **operational** perspective, e.g. *"CRM allows companies to gather customer data swiftly, identify the most valuable customers over time, and increase customer loyalty by*

*providing customized products and services"* (Rigby et al., 2002). When this category was applied to the research context, it occured that the definition defines the exact role of the team of analysts and econometric modelers, who are responsible for describing the customer with all data and attributes which are available or possible to be collected. They are collecting all the required information by using a lot of inside bank systems and then, after appropriate pre-processing they produce the results which can lead to the identification of the customer, his/her behavior, buying habits and other characteristics.

In 2005 Payne and Frow, whose CRM definition was quoted above, developed a **conteptual framework** for customer relationship management that helped broaden the understanding of CRM and its role in enhancing customer value and shareholder value as well. This study is noteworthy within the context of this thesis since it not only puts forward the typically scientific definitions of CRM, but also reveals opinions of practitioners in this respect. According to some **executives**, who had been interviewed by scientists, CRM meant direct mail, a loyalty card scheme or a database, whereas others envisioned it as a help desk, call center or an activity – populating a data warehouse or undertaking data mining. Some other business managers had considered CRM as an e-commerce solution, such as the use of a personalization engine in the Internet. According to the theorists CRM can be defined from at least **three perspectives**:

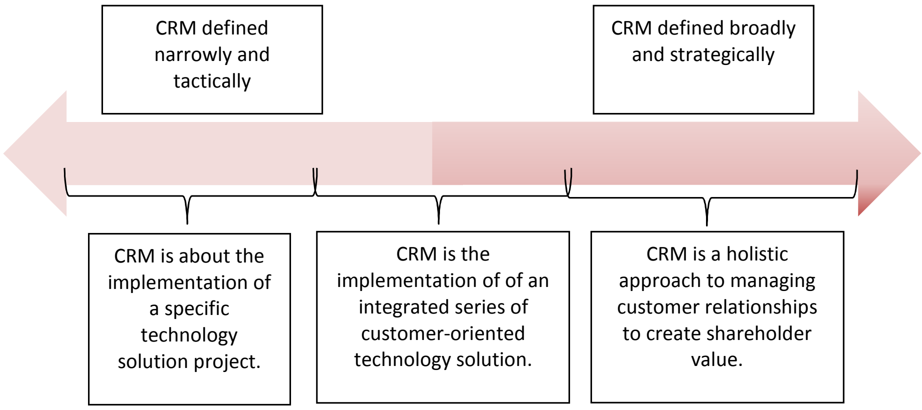
1. Narrowly and tactically as a particular technology solution (e.g. Khanna, 2001 definition) – researchers gave an example of an organization, which spent a lot of money in IT (information technology) solutions and system integrations which resulted in defining CRM in terms of its projects (especially sales force automation projects),
2. Wide-ranging technology (e.g. Stone and Woodcock, 2001 definition) – at this point researchers talked about a company which interpreted the

CRM referring to a wide range of customer oriented IT and Internet solutions,

3. Customer-centric (e.g. Buttle, 2001; Glazer, 1997; Singh and Agrawal, 2003; Swift, 2000; Zablah, Beuenger and Johnston, 2003; definitions) – according to researchers this definition is more strategic and holistic and emphasizes the selective management of customer relationship to create shareholder values.

These perspectives are presented on the picture (Figure 2.1).

*Figure 2.1 CRM continuum according to Payne and Frow research (2005)***.**




There are still many definitions of CRM available. However within the context of this thesis, the CRM definition is important not from a semantic point of view, but from translating this definition into the company`s way of working. This definition significantly affects the way the entire company accepts and practices the idea of CRM. From a strategic point of view (in the research context as well), CRM is not simply an IT solution that is used to acquire and expand a customer base. It involves a profound synthesis of strategic vision, a corporate understanding of the nature of customer value, the utilization of the appropriate information management and CRM applications, ending with the high-quality operations fulfillment and service. It was found in the banking sector that in customer-centric companies, the CRM definition is mostly realized by **the structure** of the CRM department. Each team focuses on different aspects connected with the customer and as a result it is possible to identify:

1. Operational CRM - is responsible for collecting/gathering information about customers, which can come from transactional data, customer feedback (about products, services) or other available sources,
2. Analytical CRM - is responsible for data analysis (derived from operational CRM and other sources),
3. Interactive CRM - is responsible for contacts with customers resulting from the developed marketing campaigns, the results of analytical CRM, customer needs, etc.

Analytical CRM is responsible for processing the data using different data analyses and data modeling tools. The results of the analysis and econometric models very often formulate the basis for making decisions related to the activities of the company. This link is very important because it allows the company to use of knowledge hidden in the already collected customer data.

Staying within the banking context, in the research context specifically, changing the product-centric organization into customer-centric can be achieved only if the analytical CRM methodology would include the entire organization, its employees would exchange relevant information, analysis results would be reached to all who need it, and people would act in accordance with recommendations. Because people are the most important factor linking three aspects (data, analysis and organization) of analytical CRM, it depends on people whether the data are properly collected, whether the analyses use the best available techniques and whether the results obtained, the knowledge extracted from the data is used in practice, or whether knowledge management in the company is done at the appropriate level. Therefore, in the company context, managers need to recognize that the customer-centric approach is an enterprise wide concept that requires their business to identify opportunities in order to optimize income and costs by enhancing customer value. These two effects, i.e. highlighting the customer value and optimizing the profitability, mostly by reducing the costs, create a competitive advantage resulting in higher short and long-term profitability (Bohling et al., 2006).

### 2.1.2 Cross-selling campaigns

Two key aspects mentioned earlier, which are related to the value of customer - loyalty and profitability - can be optimized through **cross-selling campaigns**, the most popular marketing solution. From customer-centric organization point of view, the most **loyal** and the most **profitable** customer is the most important and significant for the company. This optimization is realized by offering current customers some additional services and products related to previous purchases with the main goal of acquisition of a greater number of products from multiple categories (Gupta and Zeithaml, 2006). But cross-selling is also interpreted also more in-depth: it is a process, in which decisions such as assessing what products to offer, to whom and when are taken (Kamakura et al., 1991; Knott et al., 2002), and the customers are reached through multiple channels of communication: traditional mail, e-mail or offers presented during a phone call. This decision about selecting proper customers, who meet specified criteria are based on the customers` individual needs or previous behaviours (Dyche and Tech, 2001). To define it briefly (quoting a theorist) – cross selling can be defined as the implementation of relation marketing strategy (Kumar et al., 2004).

Some authors claim that cross-selling campaigns can bring many **benefits**, both to the seller and buyer. One benfit is the fact that customers with a higher number of products (in this context bank products) generate higher assets (Winer, 2001). During the customer life in the company cross-selling campaigns increase the total value of the customer (Kamakura et al, 2003). For example, Jackson (1989 in his research presented the theory that by offering six products to existing insurance policy owners during a one-year-period, using cross and up-sell campaigns, this insurance company could increase the average customer lifetime value (CLV) of an average customer by 40%. It confirms that wide usage of cross-selling is a good method for improving the business since if the well-prepared offer reaches the customer in the proper time, there is a considerable chance that the customer will benefit from it.

Cross-selling has also been associated with a higher level of customer **loyalty**. Confirmation can be found in the literature: Kamakura (2003) claimed that customers, who acquire more products from the same company find their switching costs increasing and they are more likely to stay with the firm. Such behavior results in a positive influence of the cross-selling on the relationship between these two entities, i.e. customer and company (Van den Poel and Larivière, 2004). From a practical perspective (especially in the banking sector), if the customer has got a current account and saving account in one bank, a term deposit in the second, and uses a credit card issued by a third, he/she really does not identify with any of them. If, however he/she uses the full range of services of only one bank, there is a huge chance that he/she will indentify with this bank and the next banking services he/she will benefit just from that.

Not without significance is the fact that the use of the service makes it much harder to break a contact with the bank. A good example in a bank context is a mortgage loan. If the customer, who bought such a loan, wants to leave the bank, although it is possible, it involves many burden some formalities, which usually effectively discouraged him/her from taking such action. For this type of customers the loyalty is an antecedent of cross-selling because a customer who decided to buy a mortgage loan also decided to be loyal, even if this decision was unconscious one.

Another advantage of cross-selling campaigns is that they are much cheaper than a campaign of acquiring new customers (Felvey et al., 1982). For new customers the costs of acquisition are usually large because they involve the need to conduct expensive advertising campaigns and offer special conditions to new customers in the initial period of use of services.

Cross-selling also brings benefits to customers. From a customer point of view this process refers to a customer`s propensity to make a cross-category purchase (Reinartz and Venkatesan, 2008).The first benefit is associated with the complexity of services provided to the customer. For customers it can be much more convenient to use the services of one institution because from a practical point of view, they can arrange a great number of matters in one place, which saves time. Moreover, since they are regular customers, maintaining long-term relationships with the company and buying more (Paulin et al., 1998; Ganesh et al., 2000), they can count on a special treatment and they may benefit from various discounts and rebates.

All these benefits lead to the main cross-selling **advantage**, which is maximizing the benefit of the customers who have been already won. This objective particularly gains in importance when the market in which the institution operates is saturated and it is difficult to acquire new customers. Therefore, selling products to existing customers through cross-selling strategies can increase company assets (Reichheld, 1996; Sasser et al., 1990) and increase competitive advantage. Moreover, a precise evaluation of customer profitability is a crucial element for the success of CRM (Lee and Park, 2005).

Since the acquisition of new customers has become more difficult recently and companies have already developed skills to store, share, analyze and transfer valuable information from collected data, the situation in which customer databases are exploited to the fullest is becoming more common (Fader and Hardie, 2009). That is the reason why cross-selling sometimes creates some danger. Even if there are still some groups of customers who can be selected to take a part in the cross-selling campaign, it is often better to wait and not to target.

Looking at the financial context of this research, there are also several annotations that should be adduced in order to present the daily cross-selling **threats** in the company. The observed fact is that if each group of bank products is handled by another sales branch, and each of these sections stores information about their customers in separate repositories, then the applicability of cross-selling campaigns may be confined to individual departments because of the incomplete information. It is the frequently observed situation in a product-centric organization. It is difficult to expect a satisfactory performance from such a campaign. Conducting an effective cross-selling campaign, the integration of data is required, so that customer information is as comprehensive as possible. But this operation, in addition to the challenges of technology also brings together the organizational challenges. These organizational challenges concern product managers who are afraid to start thinking about customer in complex way, as it is realized in a customer-centric organization.

## 2.2 Knowledge Discovery and Data Mining (KDDM) framework

This subchapter present the details about a theoretical framework of a process model which was taken while doing this research.

In the era of **globalization**, development of information systems and increasing ease of access to services lead to an abundance of information being presented to humanity every day. It is obvious that each person is the involuntary recipient of this information, but hardly anyone is aware that every day he or she is their own generator simply by using e-mail, ATM card, credit card or mobile phone. Most of people do not think about the fact that such simple daily actions leave **traces** ('fingerprints') of their real behavior in the shops, banks and offices. These traces, combined with today`s technological capabilities are properly collected, stored and then used to know people, their habits, needs and behaviors (Lazer et al. 2009). Companies which are selling their products and services on a large scale are the most interested in achieving **customer knowledge** from every possible and available data. When data are being collected and accumulated continuously at a dramatic pace, there is an urgent need to extract useful information (knowledge) from the rapidly growing volumes of digital data, supported by computational theories and tools. These theories and tools are the subject of the emerging field of Knowledge Discovery in Databases (KDD) (Fayad, Piatetsky-Shapiro, Smyth, 1996).

### 2.2.1 Definitions

Although data mining and Knowledge Discovery in Databases have been attracting a significant amount of researchers and industry managers for at least thirty-five years, there has been ordinary confusion in understanding the term of Knowlegde Discovery, Data Mining and Knowledge Discovery in Databases. That is the reason for explaining the published meanings of basic terms as the first step of this subsection.

**Knowledge Discovery** (KD) is the most desirable end-product of computing (Wiederhold, 1996). Moreover, this is a process which looks for new knowledge about an application domain. It consists of many steps, each aimed at completion of a particular discovery task and accomplished by the application of a discovery method (Klosgen, Zytkow, 1996). **Data Mining** (DM) includes applications, under human control, and methods, which in turn are defined as algorithms designed to analyze data or to extract patterns in specific categories from data (Klosgen, Zytkow, 1996). Data Mining is also known under many other names as knowledge extraction, information discovery, information harvesting, data archeology or data pattern processing (Fayyad, Piatetsky-Shapiro, Smyth, 1996). Then, **Knowledge Discovery in Databases (KDD)** is defined as knowledge discovery process applied to databases (Klosgen, Zytkow, 1996). According to Fayyad (1996) it is known as non-trivial process of identifying valid, novel, potentially useful and ultimately understandable pattern in data. This definition is the most popular as it was developed by revising the original definition published in 1991 by Frawley (et al.). Finally, putting together, according to Klosgen and Zytkow (1996) **Konwledge Discovery and Data Mining** (KDDM) concerns the Knowledge Discovery process applied to any data source. Moreover, according to two independent sources (Reinartz, 2002; Cios, Kurgan, 2005) KDDM has been defined as the most appropriate name for the overall process of KD. More particularly, KDDM includes the entire knowledge extraction process, starting from how the data is stored and accessed, how to develop efficient and appropriate algorithms, which can be successfully used to analyze massive datasets, how to interpret and visualize the results, and how to model and support the interaction between human and machine (Fayyad, Piatetsky-Shapiro, Smyth, 1996). KDDM supports learning and analyzing the application domain and DM is always present as one of the steps in the process.

### 2.2.2 View on history and motivation

The first concept of the KDDM process model started its history in a workshop on KDD in the **early 1990s** (Piatetsky-Shapiro, 1991). The main reason to define the model was the acknowledgement of the fact that knowledge is the end product of data-driven discovery process. One of the key findings of the workshop was also the recognition of the need to develop interactive systems that would provide visual and perceptual tools for data analysis. After this workshops, KDD community was developing the idea of the process beginning from single DM techniques such as decision tree, clustering algorithm, with a very small support for the overall process framework. Such systems were useful for the researchers who had understood DM techniques (Zytkow, Baker, 1991; Klosgen, 1992; Piatetsky-Shapiro, 1992; Ziarko, 1993). Because of the lack of DM methods and little attention focused on the support of layman analysis, the first KD modeling systems had minimal commercial success (Anand, Branchman, 1996). In 1996 group of researchers (Fayyad, Piatetsky-Shapiro, Smyth, Uthurusamy) published *"Advances in Knowledge Discovery and Data Mining"* in which they presented in this publication a process model that resulted from the interactions between researchers and industrial data analysis. The new driver was the fact, that the authors focused on providing the support for the complicated process of highly iterative knowledge generation and showing the close involvement of a human analyst in the majority of steps instead of addressing the existing or new DM method. The research presented in the book evolved with two major types of process models: the human-centric model, which emphasized the interactive involvement of a data analyst during the process and the data-centric model that emphasized the iterative and interactive nature of the data analysis tasks (Fayyad, Pietetsky-Shapiro, Smyth, Uthurusamy, 1996). But the both models of all two mentioned types have something in common. They treat the process as a highly interactive, with a high saturation of complexity. Moreover, they recommend that KDDM process may use, or at least should consider the use of a set of DM methods, while admitting that the DM stage constitutes only a small portion of the overall process (Anand, Branchman, 1996; Fayyad, Piatesky-Shapiro, Smyth, Uthurusamy, 1996). To learn more about the differences and similarities of the human-centric and data-centric models please consult the mentioned literature.

In 2006, the researchers Kurgan and Musilek went to a great lenght to consolidate in one article the historical overview, description and future directions concerning the standard for the KDDM process model. They also presented a comprehensive comparison of several process models and discussions on both academic and industrial applications. Therefore, according to them, there are four main factors why the KDDM was formally structured as a process. The first factor is the result of an observation of the problems associated with a simple application of DM methods to input data. Statistical literature often called it ‘data dredging’. Such an approach can lead to **meaningless conclusions** (Fayyad, Piatetesky-Shapiro, Smyth, 1996). Thus, based on these observation, before the KDDM process is used, it should be preceded by an investigation, which predicts if the end product will be useful to the recipients (users) (Fayyad, Piatetsky-Shapiro, Smyth, 1996d). Moreover, according to Kurgan and Musilek (2006), it is important that only well-defined and formal development methods let achieve desirable properties with the success.

Another motivator to structure the KDDM methods is connected with the proper understanding of the process and of the end-user needs as well. Very often it is easy to observe very typical **human behaviour** associated with the knowledge-searching tasks connected with large amounts of untapped and potentially valuable data. People are usunally not willing to dedicate time and resources toward formal methods of knowledge seeking but rather rely on data experts as a source of valuable information (Rouse, 2002). They do this because they very often **feel uncertain** about an unknown field of new technology and processes that need to be applied to provide an appropriate solution (Rouse, 2002). That is the reason of heavy need of standardization of developed solutions (Kurgan, Musilek, 2006).
The third factor, which is also very important from this research context, is associated with providing support for management problems. In some fields it is very common that KDDM projects involved a relatively large number of people working in one team, so a detailed schedule is a basic need. In order

to prepare such a plan and its milestones project management specialists need a **clear definition** of KD and DM, about their contents and how to carry out the process out. The project specialists usually try to define each milestone as a concrete, specific, measurable event used to define completion of particular phases of the overall projects (Brooks, 1995). Thus, they could be properly defined only in the context of the well-defined larger framework of the process (Kurgan, Musilek, 2006).

Finally, the last important factor - KDDM process standardization - is needed in order to provide **an unified view on existing process description** and to allow an appropriate usage of technology to solve a specific business problem in practice (Reinartz, 2002).

### 2.2.3 The KDDM process models

A process model is the set of tasks to be performed to develop a particular element, as well as output (elements that are produced in each task) and inputs (elements that are necessary to do a task) (Pressman, 2005). The goal of the process model is to make the process repeatable, manageable and measurable (Marbán, Mariscal, Segovia, 2009).

A **KDDM process model** consists of a set of processing steps to be followed by practitioners while executing the project and it describes procedures that are performed in each step to start, go through and finish the project. The basic process model was proposed by Fayyad, et al. in 1996. Since then several different KDDM process model have been developed and introduced in both academic and industrial fields. In the beginning there was a rush to develop DM algorithms that were capable of solving all the problems of seeking the knowledge in the data. Besides, tools were also developed to simplify the application of DM algorithms. Finally, the year 2000 noted the most important milestone from the perspective of KDDM process models: CRISP-DM (CRoss Industry Standard Process for DM) was published

(Chapman et al., 2000). This process model is the most used methodology for developing KDDM projects.

All the KDDM process models consist of multiple steps executed in a sequence, which often includes loops and interactions. Each subsequent step is initiated upon the successful completion of a previous step and requires a result generated by the previous step as its input. Model activities range from the task of understanding the project domain and data, through data preparation and analysis, to evaluation, understanding and application of generated results. The main difference between the KDDM process models is the proposed number and scope of their specific steps. Moreover, compared with academic projects, industrial ones are usually concerned with different types of data, have more complex application scenarios and are associated with different burdens and pitfalls (Kurgan, Musilek, 2006).

### 2.2.4 Cross-Industry Standard Model for Data Mining (CRISP-DM)

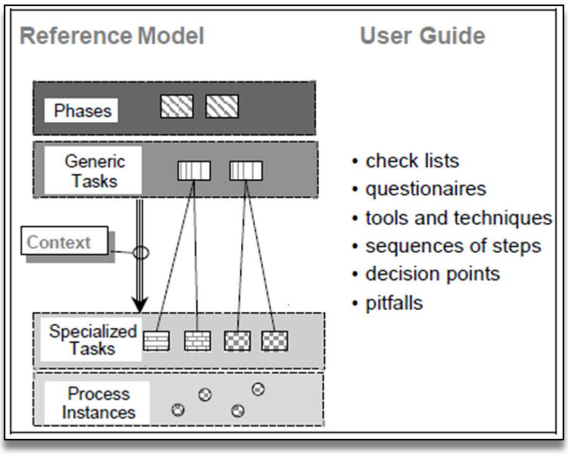
**Data mining** aims to extract knowledge and insight through the analysis of large amounts of data using sophisticated modeling techniques. It converts data into knowledge and actionable information. In general, the data to be analyzed may reside in well-organized data marts and data warehouses or may be extracted from various unstructured data sources. A data mining procedure has many stages. It typically involves extensive data management before the application of a statistical or machine learning algorithm and the development of an appropriate model. Data mining models consist of a set of rules, equations, or complex transfer functions, that can be used to identify useful data patterns, understand, and predict behaviors. According to the literature (e.g. Larose, 2005; Maimon and Rokach et al., 2005; Witten and Frank, 2005), they can be grouped into two main classes according to their goal, as follows:

1. **Supervised/predictive** models - the goal is to predict an event or estimate the values of a continuous numeric attribute. In these models there are input fields or attributes and an output or target field. It is possible to define:
   i. **Classification** or propensity models – where the target groups or classes are known from the start. The goal is to classify the cases into these predefined groups; in other words, to predict an event. The generated model can be used as a scoring engine for assigning new cases to the predefined classes. It also estimates a propensity score for each case. The propensity score denotes the likelihood of occurrence of the target group or event.
   ii. **Estimation** models - similar to classification models but with one major difference. They are used to predict the value of a continuous field based on the observed values of the input attributes.
2. **Unsupervised/undirected** models - there is no output field, just inputs. The pattern recognition is undirected; it is not guided by a specific target attribute. The goal is to uncover data patterns in the set of input fields. Unsupervised models include:
   i. **Cluster** models - the groups are not known in advance, so the algorithms are needed to analyze the input data patterns and identify the natural groupings of records or cases. When new cases are scored by the generated cluster model they are assigned to one of the revealed clusters.
   ii. **Association and sequence** models - do not involve direct prediction of a single field. In fact, all the fields involved have a double role, since they act as inputs and outputs at the same time. Association models detect associations between discrete events, products, or attributes. Sequence models detect associations over time.

When properly built, all these models can identify the right customers to contact and lead to campaign lists with increased density/frequency of target customers. They outperform random selections as well as predictions based on business rules and personal intuition.

DM is a creative process which requires a number of different skills and knowledge coming from DM specialists, which means that while realizing such project in the industry, there is a need to define a standard approach which will help translate business problems into data mining tasks, suggesting appropriate data transformations and DM techniques. The CRISP-DM process, that is industry-oriented, addresses mentioned attributes by providing a framework for carrying out data mining projects. This model was developed by a consortium of leading data mining users and suppliers: DaimlerChrysler, SPSS, NCR and OHRA (Chapman et al., 2000). The CRISP-DM methodology is described in terms of a hierarchical process model, comprising four levels of abstraction, from general to specific: phases, generic tasks, specialized tasks and process instances. It is illustrated in Figure.

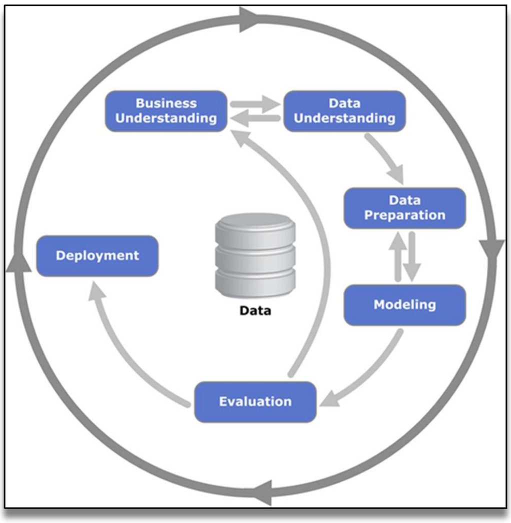
*Figure 2.2.4.1 The Level Breakdown of CRISP-DM methodology (Chapman et al., 2000).*



At the top of the level the CRISP-DM process is organized into a small number of **phases**. Each phase consists of several second-level **generic tasks**. Another level, the **specialized task** level is the place to describe how actions in the generic tasks should be carried out in specific situations. The last one **- the process instance** level is a record of actions, results of and actual DM engagement.

CRISP-DM is divided into six phases which are carried out in the KDDM project (Figure 2.2.4.2).

*Figure 2.2.4.2 The CRISP-DM process model (Chapman et al., 2000).*




The phases are described by Marban et al (2009):

1. **Business understanding** – focuses on understanding the project objectives and requirements from a business perspective, then converts this knowledge into a DM problem definition and a preliminary plan designes to achieve the objectives.
2. **Data understanding** - considers the data requirements for properly addressing the defined goal and an investigation of the availability of the required data. This phase also includes initial data collection and

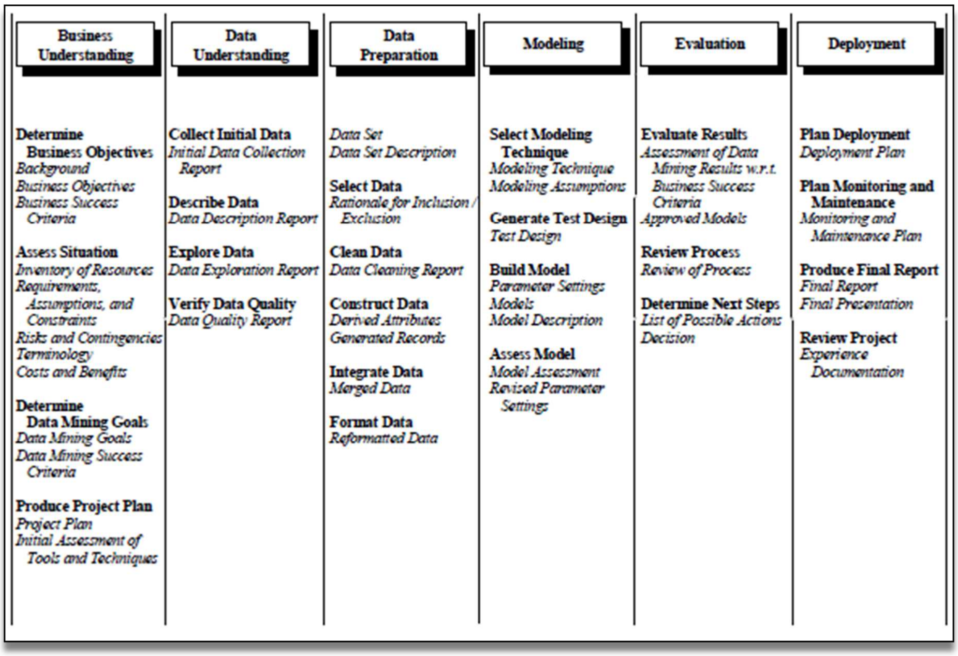
exploration with summary statistics and visualization tools to understand the data and identify potential problems in availability and quality.

3. **Data preparation** - involves the acquisition, integration, and formatting of the data according to the needs of the project. The consolidated data should then be cleaned and properly transformed according to the requirements of the algorithm to be applied. New fields such as sums, averages, ratios, flags, and so on should be derived from the raw fields to enrich customer information, to better summarize customer characteristics, and therefore to enhance the performance of the models. This phase is likely to be performed repeatedly and not in any prescribed order.
4. **Modeling** - involves the examination of alternative modeling algorithms and parameter settings and a comparison of their fit and performance in order to find the one that yields the best results. Based on an initial evaluation of the model results, the model settings can be revised and fine tuned.
5. **Evaluation** - The generated models are then formally evaluated not only in terms of technical measures but also, more importantly, in the context of the business success criteria set out in the business understanding phase. At the end of this phase, a decision should be reached on how to use of the DM results.
6. **Deployment** – Model construction is generally not the end of the project. Even the best model will turn out to be a business failure if its results are not deployed and integrated into the organization’s everyday marketing operations. A procedure should be designed and developed to enable the scoring of customers and the updating of the results. The deployment procedure should also enable the distribution of the model results throughout the enterprise Finally, a maintenance plan should be designed and the whole process should be reviewed.

Lessons learned should be taken into account and the next steps should be planned.

The phases above present strong dependencies and the outcomes of a phase may lead to revisiting and reviewing the results of preceding phases. The nature of the process is cyclical since the data mining itself is a never-ending journey and quest, demanding continuous reassessment and updating of completed tasks in the context of a rapidly changing business environment. Figure 2.2.4.3 outlines the phases and generic tasks that CRISP-DM proposes to develop a DM project.

*Figure 2.2.4.3 The CRISP-DM phases and generic tasks (Chapman et al., 2000).*




The process of changing the product-centric into the customer-centric company, starting with the changing the marketing campaigns from being single product to multi-product is based on the KDDM process, specifically on CRISP-DM process and this project includes six mentioned phases. In the next chapters the reader will get the details of the whole project.

# 3. A methodology for multi-product offering marketing campaign

## 3.1 The overview

Customer propensity-to-buy and customer segmentation problems in marketing field have been handled previously many times. However, in this study there is a new empirical design of usage of popular methods proposed. The empirical design consists of three steps:

1. **Scoring** - estimating the propensity-to-buy scores as a result of the propensity-to-buy models built for four bank products,
2. **Segmentation** – grouping the customers into the homogeneous clusters,
3. **Product set selection** – assigning the set of the most appropriate bank products to every cluster.

This multi-product offering (henceforth reffered to **MPO**) design is performed in order to identify the most likely to buy customers, to identify the most popular and useful products which could form specific packages, with the aim of implementing the strategies to **manage the customers**. The proposed approach will be tested with an appropriate test and learn strategy. Chapter 3 describes the scoring, segmentation, product set selection and test and learn strategy framework.

## 3.2 Scoring

In this section a detailed overview of the estimation of the propensity-to-buy scores are described as the realization of the Modeling stage of CRISP-DM framework.

The **regression** modeling is defined as a functional relationship between Y and X, which gives an idea about the otherwise non-deterministic Y, where X is a set of explanatory or independent variables called $X_1, X_2, \ldots, X_k$ and Y is the response or dependent variable:

$$Y = f(X_1, X_2, \ldots, X_k) \qquad (3.1.)$$

Knowledge of Y variable is crucial for decision making but it is not deterministic. Moreover, X is available at the time of decision making and it is related to Y. An explanatory variable could be numerical: discrete (e.g. number of bank cards in customer portfolio) and continuous (e.g. deposit balance) or categorical: ordinal (e.g. income group – high/medium/low) and nominal (e.g. gender – male/female). A dependent variable could be continuous (e.g. total amount of credit the customer wants to buy) and discrete (e.g. number of products which can be bought) and binary (e.g. whether the customer would default on payment or not 1/0). The type of data mining method which should be used to modeling mainly depends on dependent variable Y: if Y is a continuous variable ordinary least squares regression is used, and when Y is binary (0 or 1), a logistic regression is recommended.

When building **propensity-to-buy models**, the occurrence of a customer`s wish to buy a given bank product (1) or not to buy it (0) is being modelled. To put it precisely, this method (**logistic regression**) models the logarithm of the odds of the event occurring as a linear function of a set of covariates. The training set of observation contains data divided on independent variables (X) and one target, dependent binary variable (Y). Under the assumption that $Y_i$ and $Y_j$ are independent for all $i \neq j$, using maximum likelihood as the method of estimation, having set of parameters: $\alpha$, $\beta_1$, $\beta_2$,...,$\beta_k$ to be estimated it is necessary to define the model as follows:

$$P_i = Prob(Y_i|X_j) = \frac{1}{1+e^{-L_i}} \tag{3.2}$$

where

$$L_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} \tag{3.3}$$

The propensity-to-buy modeling is a typical analytical case study in the company which can be realized using the stages of KDDM process, CRISP-DM process specifically. It involves two sides of process basic company entities: product managers, who represents **business** part and statistics experts, who represents **analytical** part. Because the study concerns the real case study and is connected with some consequences, the stages of CRISP-DM process are determined by the rules and procedures which were applied in the company.

From the bank perspective, the process of customer propensity-to-buy (for a given product) modeling may include 6 important points:

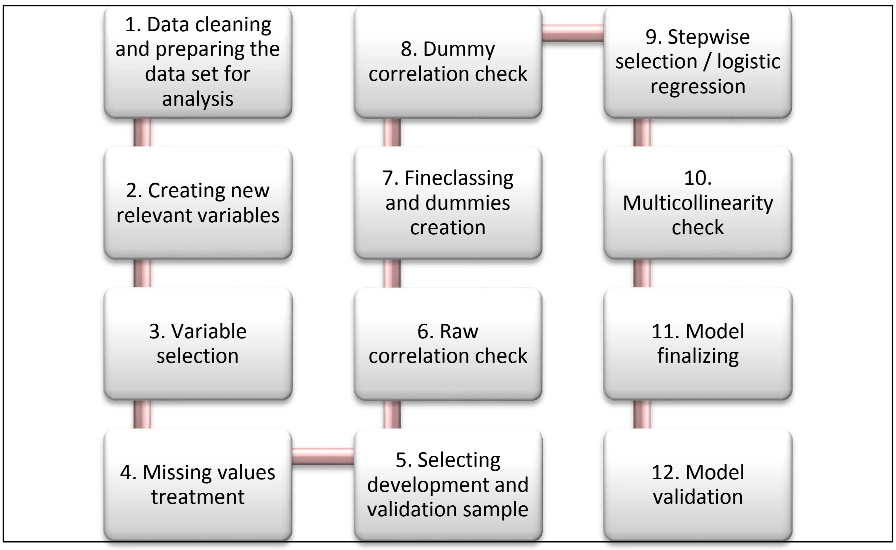
1. First it is necessary to formulate the business problem as a statistical problem. This can be simplified by specifying the objective function and target/dependent variable. It is the business understanding stage.
2. Data exploration and description with data quality, which formulates the data understanding step.
3. The most time consuming step – preparing the data required for model building. It is just completing the X-set, set of independent variables. This is synonymous for ther data preparation stage.
4. Developing the model with all the points, according to modeling algorithm. The modeling algorithm is followed in the correspondence with the bank experts` modeling rules and the author does not have the complete freedom about the implementation details. All the steps which describe the modeling algorithm will be discussed in the next

part of this chapter. This step is the same as the Modeling stage in the CRISP-DM.

5. Launching the model results into the real campaign. This part could be regarded as the evaluation step of the model.
6. Model validation and monitoring in order to avoid overlearning of the model and identification the limitation of the model. This can be summarized as the deployment step.

The implementation of modeling algorithm is crucial to generate the propensity-to-buy scores. It is considered as a single stage of entire CRISP-DM framework, although it consists of twelve steps which are illustrated in Figure 3.2.1. Since these steps were decided in correspondence with bank experts and are in some way specific for this type of organization (the given bank), some details about each are discussed below in order clarify their rules.

*Figure 3.2.1 Propensity-to-buy modeling algorithm.*



**1. Preparation of data.**

This is the first step in most data mining processes since the customer data is stored on the servers in different tables. Then it is possible to merge, divide, union, join table with different data from these different tables and, as a result, we get merged data which we clean using all available cleaning methods. In the final data set there are data which are aggregated to customer level, ready for analysis. During the cleaning process the erroneous values have to be identiefied and inconsistency in the values of variables has to be checked .

**2. Creating new variables.**

Available data can be divided into some logical groups: demographic, socioeconomic, product level and behavioral. New relevant variables, if necessary (and in most cases they are necessary) are to be created from the existing ones. For example – utilization is a derived variable which is created from balance and credit limit.

**3. Variable selection.**

During this part the focus lies on some clue rules: selection of variables depends on the purpose of modeling, irrelevant variables are to be dropped, and variables with a large percent of missing values are also to be dropped. Note that, there is no limit to the number of variables to be selected for analysis. And the last rule is about variables which are not relevant for the purpose of modeling. If so, such variables are not to be considered in the process.

**4. Missing values treatment.**

The most intuitive treatment of missing data in the academic data sets is the deletion of all cases with missing values. However, this common approach is unuseful in the real data world, where in practice a missing value is also a value for the manager and researcher. This statement is supported by the finding that deleting the cases with missing values could bias any conclusions about the population drawn from the collected sample (Hair et al., 2006, King et al., 2001). In 1976 Rubin developed the classification of missing values explaining why deleting cases with missing values should be avoided. According to this partition there are three types of missing data which can occur: (i) data which are not missing at random (NMAR), (ii) data which are missing at random (MAR) and (iii) data which are missing completely at random (MCAR). In the real data set, in the research context, in all cases when missing values are not completely at random, deleting records with missing entries not only reduces the sample size, but also results in biased conclusions because very often missing values means that customer merely is more reluctant to give the answer or leave the information about him/herself. Therefore, it is more advisable to use specific imputation methods to tackle missing values instead of deleting cases with missing values. Very often, multiple imputation and model-based imputation methods are preferable, since they do not lower the sample size and they are unbiased. The multiple imputation methods produce multiple imputed versions of the original data set and each version contains different values for the missing entries, which represents the uncertainty of the imputed values. The different results of each imputed data needs to be combined into a single statistical result. Another imputation method which is preferred is model-based imputation. It involves maximum likelihood estimation of the underlying process which generates the missing values.

In this thesis the following approach to tackle missing data is recommended. Firstly, the level of missing data per variable is calculated. According to the rule of thumb, variables with more than 20% missing data are considered for deletion. These values should not be used for modeling unless the fact of missing value has a special significance and can be replaced by some meaningful number. Finally it is verified that the missing data are MAR or MCAR by means of Little`s MCAR test. In general, missing value treatment depends on the percentage of missing values. When missing values are:

- less than 1% of data set - the observations were deleted or multiple/single imputation was used
- 1-20% multiple imputation or model based imputation is used.

Irrespective of the type of missing data, the missing values are imputed with SAS Base 9.1. tool (proc mi, proc mianalyze).

**5. Selecting development and validation sample.**

After the data set preparation, it was divided into a development sample, to build and learn the model, and validate the sample to check if the model knows how to estimate the probability of response correctly and to check if it is not over-learned. The most common division is 60/40 proportion, what means that the development sample represents 60 percent of the data set and the validation sample is the remaining 40 percent. In practice it is easier to split the data set into part of target=1, called 'good' and part of target=0, called 'bad'. Then each part is split randomly in 60-40 samples. A combination of proper parts of 'goods' and 'bads' will result in development and validation samples. While having these two samples it indicates whether the model is robust or not.

**6. Raw Correlation Check.**

Correlation is the degree of association between two variables. A correlation check helps analysts to eliminate multicollinearity from the system. A set of independent or explanatory variables are said to have 'Multicollinearity' if there is any linear relation between them. The correlation coefficient ($R^2$) indicates the degree of association between two variables ($0<=R^2<=1$). However, correlation among two variables does not necessarily imply high multicollinearity but very high correlation (>0,9) does imply multi-collinearity. When correlation among the raw variables is checked, in most cases it provides analysts a tool to significantly reduce multicollinearity among variables.

**7. Fineclassing and dummies creation.**

The seventh part of the model building processes is fineclassing, which determines which characteristics are worth of consideration in the model development. Each characteristic is investigated to determine 'good' (target=1) or 'bad '(target=0) trends at the attribute level. Once trends have been identified, attributes are grouped together to smooth out fluctuations in data and they are called dummies. Then definitions are imputed and read on the meaning of log-odds (weight of evidence), which determines and compares proportion of 'goods' in the attribute with the proportion of 'bads' in the attribute. A value of '0' implies equal proportion of 'goods' and 'bads' in an attribute. A positive value indicated proportionally less 'bads' then 'goods'. Magnitude of the value indicates by how much less. With negative value is the opposite case. In the results report of fineclassing there is also Information Value available, which measure how well the characteristics can determinate between 'good' and 'bad' and whether it should be considered for modeling. When Information value is <0.03, the characteristic is not predictive and should not be considered for modeling. In case of an Information value within the range in 0.03-0.1 the characteristic is predictive and should be considered for modeling. When the information value is >0.1 then the characteristic is very predictive and should definitely be used in modeling. Grouping of attributes is required to smooth out reversals in trend and to combine attributes with similar characteristics.

**8. Dummy correlation check.**

Fineclassing helps in creating dummies and once dummies are created we need to run the correlation check on these dummies, which is the next step. This is performed in order to take care of any significant multicollinearity effects that may exist among the dummies. The most popular cut-off used for dummy correlation check is 0.5 while for raw correlation cut-off=0.9 is used mostly. There are some advantages of using dummy variables:

- they are easy to tackle outliers and influencial points
- provide a very good way of tackling situations where a particular variable does not follow a trend
- give equal weight to observations that behave in the same way
- provide a good means of addressing the issue of a large number of missing values.

On the other hand, there are also some disadvantages of using dummy variables available:

- the number of unique scores is very low and it leads to problems in case of deciding cut-offs
- very similar values can be assigned radically different weights
- there is problem of interpretation – it is not easy to comprehend as original variables.

**9. Stepwise selection.**

Finally, in this stage a logistic regression algorithm can be used to investigate the relationship between endogenic variable and set of describing variables.

Procedure in SAS Base 9.1. tool, which was used to fit conditional logistic regression models for binary response data, operates with the method of maximum likelihood. The maximum likelihood estimation is carried out with Fisher scoring algorithm, and link function is captured by logit function. Fisher scoring algorithm is equivalent to fitting by iteratively reweighted least squares and it is based on the expected information matrix. In the case of a binary logit model, the observed and expected information matrices are identical, resulting in identical estimated covariance matrices. With logistic regression estimation five effect-selections methods are available, although in this study for every model that was built, a stepwise selection method was used (Hosmer, Lemeshow, 2000).

**10. Multicollinearity check.**

Multicollinearity, which is a set of independent or explanatory variables with any linear relation between them, is used to identify problem of parameter estimated unreliability. To detect multicollinearity it is necessary to measure the Variance Inflation factor

$$VIF = \frac{1}{1-R^2} \tag{3.4}$$

and Condition Index

$$CI = \sqrt{\frac{\max(eigen\ value)}{individual\ eigen\ value}} \tag{3.5}$$

After detection of multicollinearity we should remove it by looking into Variance proportions table for the row with highest CI, then identify variables with highest factor loadings in the row and drop the variable which is least significant. Main rule in removing multicollinearity is

$$VIF > 1.75 \equiv> Multicollinearity \tag{3.6}$$

**11. Model finalizing.**

The model on the development sample should be finalized on the basis of the model statistics:

i. VIF indicates the degree of multicollinearity.
ii. Chi-square value for each explanatory variable: the chi-square value indicates the level of significance, for example – the impact of an explanatory variable on the dependent variable
iii. Concordance – among all pairs formed from the 0 and 1 observation of the dependent variable, the percentage of pairs where the probability assigned to an observation with the value 1 for the dependent variable is greater than that assigned to and observation with value 0. Percentage of concordant pairs should be at least greater than 60.
iv. Rank ordering:
    - Order data in descending order of predicted values
    - Break into 10 groups
    - Check if average of actual is in the same order as average predicted.
v. Kolmogorov Smirnov test (KS statistics) – defined as the absolute difference between cumulative percentage of ‘goods’ and cumulative percentage of ‘bads’ (max KS should be > 20)
vi. Lift Curve (Lorenz Curve) – indicates the lift provided by the model over random selection.

**12. Model validation.**

When the model is finalized it is necessary to prepare model validation on the validation (hold-out) sample. It consists of validation rerun step and second one – scoring the validation sample. Validation rerun means rerunning the model on the validation sample, checking the chi-square values, the level significances and p-values for each explanatory variable. The p-values should not change from the development sample to the validation sample. Rank ordering should also be similar. Scoring the validation sample is done by using the parameter estimated obtained from the development sample. After this step rank ordering should also be checked.

In summary, propensity-to-buy modeling process, with its all described above modeling steps, lead to the final status that every bank customer is assigned an appropriate score that estimates his/her willingness to buy a given bank product. All models, which are being operated in the bank were created in accordance with the action methodology described above.

## 3.3 Segmentation

Segmentation, meaning cluster analysis, as a statistical method is a tool for exploratory data analysis and its aim is to arrange the objects into groups in such a way that the degree of similarity between objects belonging to the same group was the largest, and the objects of other groups as small as possible. To put it more simply, objects in a given cluster tend to be similar to each other in some sense, and the objects in other clusters tend to be dissimilar. That is the reason why cluster analysis can be used to detect structures in data without outputting interpretation and explanation. Briefly, cluster analysis only detects structure in data, without explaining why they occur. By analyzing the differentiating variables it is possible to find the reason of existing of each group. In order to find five disjoint clusters of observations (meaning customers) k-means method was applied to coordinate data. The main reason for this choice was the enormous number of data sets. From the perspective of the calculations, this method can be regarded as a 'reversal' of analysis of variance (ANOVA). It starts with *k* random clusters and then objects are moved between those clusters with a view to minimizing variability within clusters and maximizing variability between clusters. In other words, the members of the same group are characterized by the maximum similarity and the members from the other groups are characterizes by the minimum likeliness. This is the 'inverse' analysis of variance analysis in the sense that the significance test in ANOVA compares the variability between the intra-group variability in carrying out the significance test for the hypothesis that the mean in groups do not differ from each other. From the mathematical point of view the purpose of k-means algorithm is to assign the *n*-dimensional data vectors to code vectors $r_i$ where $i \in [1, N]$, with the least average quantization error. The average quantization error is given by formula:

$$D = \frac{1}{k} \sum_{i=1}^{k} d(x_i, r) \tag{3.3.1}$$

where $k$ is the number of $x_i$ elements which are assigned to the code vector $r$ and $d$ is the measure of the quantization error and it is mostly square error determines for the *n*-dimensional vectors as:

$$d(x, r) - \sum_{j=1}^{n} (x_j - r_j)^2 \tag{3.3.2}$$

The k-means algorithm can be described in several steps. First step chooses $N$ code vectors and specifies the maximum quantization error $e$. Algorithm starts with $m=0$ iteration. Average quantization error in each $m$ iteration is defines as $D_m=\infty$. Then, group of $M$ data vectors are divided on $N$ groups. Vector $x_j$, where $(j \in [1, M])$ is assigned to given $i$ th group, if and only if the inequality is true for all $r_k$ different from $r_i$:

$$d(x_j, r_i) \le d(x_j, r_k) \tag{3.3.3}$$

For these assumptions average quantization error is defines as:

$$D_M = \frac{1}{M} \sum_{i=1}^{M} d(x_i, r) \tag{3.3.4}$$

and it is calculated for code vector $r$, which comes from this group that data vector $x_i$ has been classified to. All groups of vectors were determined with centroids and all code vectors (*rj*) were assigned to centroids. If below inequality:

$$\frac{D_{m-1}-D_m}{D_m} < e \qquad (3.3.5)$$

is met, the algorithm ends. Otherwise, value of *m* should increase and each step should be repeated. The k-means algorithm continuously adjusts the code vectors to existing data and if it is necessary, wrongly classified data vectors are moved to other groups.

## 3.4 Multi-product offering (called MPO) idea

This dissertation considers the one particular medium sized **universal retail bank**, with foreign capital, placed in the East-central Europe. The cross-selling campaign process is supported by four propensity-to-buy models which are used to prepare single product cross-selling campaigns for four groups of bank products: Credit Product no1 (called in the thesis CP1), Credit Product no2 (called in the thesis CP2), Credit Product no3 (called in the thesis CP3) and Deposit Product no1 (called in the thesis DP1). In the beginning of each marketing campaign, there is a large customer database available to use. In the common single product offer approach product manager decides that in the first stage customers for Credit Product no1 offer are selected. So next, analysts and modelers are supposed to implement and then use the propensity-to-buy model and choose the customers who are the most likely to buy this product (i.e. the customers, who have got the highest propensity-to-buy scores). Then, after excluding selected customers, each step is repeated for Credit Product no2 (imposing the proper propensity-to-buy model, choosing the most likely to buy customers and leaving the rest of the database to other products). Afterwards, the same steps are taken for Credit Product no3 and for Deposit Product no1. There are some disadvantages which can be seen in this process. First, if customer database is quite constant in number, which is very possible in today`s market full of competitors and after the financial crisis, this customer database could have the end, it could be exploited to the fullest and the customers` portfolios could finally reach their maximum of bank product saturations. Second, it is always a threat that if customer propensity-to-buy score is the same in every n-campaign the selected customers to be offered are the same. The third threat affects the customers who are likely to buy all of the products or those who want to buy just more than one. In that case, when the order of selecting group of people for given product is always the same, there is no chance to change the offer for given customer. For example, if there is a group of customers who are likely to buy all of the proposed bank products, this group will be unfortunately offered still with the same and only one Credit Product no 1 because the bank prepares only single product offer and creating the customer database process for a given campaign starts always from the same product (Credit Product no 1). In the result, very often it is necessary to deal with the situation which is illustrated in Figure 3.4.1.

*Figure 3.4.1 Customer division in the product-centric organization.*


Finally, the single-product offering process assumes that in fact the product is this subject, which is the most important in marketing offering, just like in the product-centric organization. To change it to a customer-centric one it needs to change the subject into the customer instead of the product.

The offering, which is based on the segmentation created from scores of propensity-to-buy models is a solution for the threats which are mentioned above. Preparing the customer segmentation means that there is a solution in forming the groups of customers who are similar to each other and simultaneously different between the groups. The idea could be seen as a remedy for the situation, where each product manager tries to have the best, from his/her point of view and from the product perspective, customers. In such an approach managers are able to find customers who are likely to buy all of the products and also others, who are not likely to buy anything. Customers are divided according to the propensity-to-buy scores which are cumulated into the deciles to keep some kind of score standardization. Based on the propensity-to-buy deciles it is possible to distinguish several (in this research five) clusters of similar customers and prepare the most suitable offer for each cluster. The offer could be single offer, as it used to be, or it could consist of some products instead of only one, as customers can choose which one is the best, which one is the most appropriate for him/her in given moment or even can buy all of proposed products. The MPO seems to be more comfortable for customer as it tries to meet all of the customer`s likeliness to buy expectations since the offer is the direct answer for the propensity-to-buy scores which are present in the given segment (cluster). Figure below is the reflection of the situarion described.

*Figure 3.4.2 Customers division in the customer-centric organization, based on the multi-product offering according to the segmentation of the propensity-to-buy scores.*

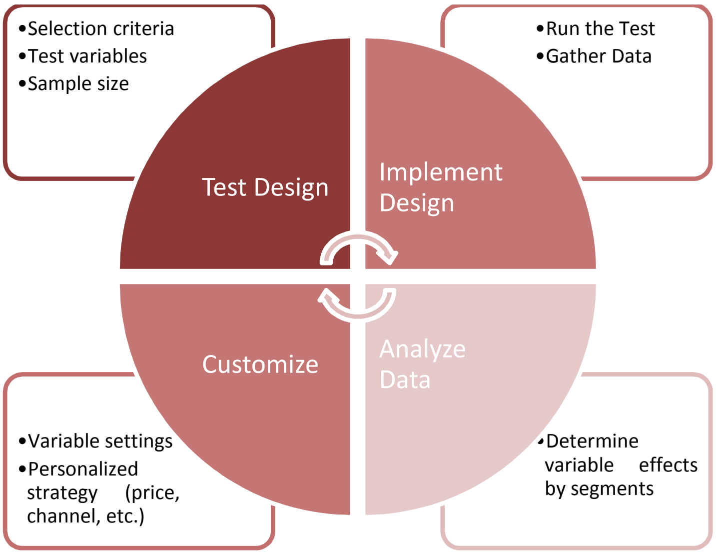

.

## 3.5 Test and learn strategy

The new proposed approach has to **be tested** before it will be used regularly during the marketing campaigns. In this chapter assumptions of the MPO offering test and strategy are presented. In the Figure 3.5.1 there steps of **test and learn strategy**[2] are presented.

[2] Test and Learn strategy - is a set of practices followed by retailers, banks and other consumer-focused companies to test ideas in a small number of locations or customers to predict impact. This strategy has been systematically applied as far back as 1988 by Capital One. Capital One has been aggressive about testing since the firm was founded, testing everything from product design to marketing to customer selection to collection policies (Davenport and Thomas, 2009; Fishman, 2009; Fleenor, 2009; Angrisani, 2009; Wong, 2009).

*Figure 3.5.1 Test and learn strategy.*




The first task of testing is a trial to think about and find the answers on key elements of the design part below:

i. Which customers are necessary to consider?
ii. How many test groups should exist?
iii. How to construct the metric to compare the groups?
iv. How many customers should be in each group?

The assumption about selection criteria from **Test Design** part is the first key element of this step. In the case described, a group of customers is selected from the entire available bank customer databases, but it is particularly important that only the customers with all four scores of the created propensity-to-buy models and with all five segments assigned sgould be taken into account.

Then, it is necessary to complete the following stages of **Test Design process**:

i. Define business objective to be tested.
ii. Restate objective as a statistical hypothesis.

iii. Define the groups and control groups.
iv. Identify sample size constraints and perform sample sizes analysis to define minimum samples size to be tested in each group.

In order to specify the **business objective** there is a need to look for the answer to the question what the analysts and managers really do need to find out from the test. In this thesis the main business objective is to check, if the MPO approach gives the same or better response rates results comparing it with so far single-product offering approach. This answer can also specify which type of organizational norm (a product-centric or proposed customer-centric) let company achieve better results. Response rates are analyzed for each propensity-to-buy model and for each created segment. Customers from **Test (Champion) groups**[3] are proposed with MPO, customers from **Control (Challenger) groups**[4] are proposed with single-product offer. This is translation of product managers` need. But there is also a second objective – **analytical perspective**, which is about checking the correctness of created segments. To meet this objective appropriate Control groups are predicted to be launched as well. Customers from such clusters will be given marketing offers from other segments. When the business objective is precisely defined it is easier to go through further steps correctly. In the next step a statistical point of view is presented. Hypotheses are formulated and are shown in the Table 3.5.1.

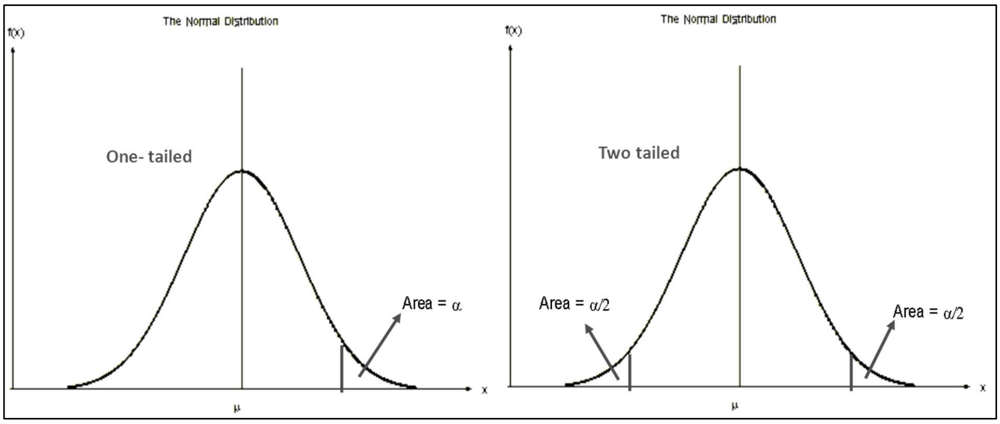
*Table 3.5.1 Formulate statistical hypothesis.*

| Hypothesis type | Null hypothesis | Alternative hypothesis |
|---|---|---|
| Two-tailed | $RR_1-RR_2=0$ | $RR_1-RR_2\neq0$ |
| One-tailed | $RR_1-RR_2\geq0$ | $RR_1-RR_2<0$ |
| One-tailed | $RR_1-RR_2\leq0$ | $RR_1-RR_2>0$ |

[3] Test group is a group where experimental assumptions/treatment are going to be applied and checked.
[4] Control group is comparative group to test group, which is applied with standard or any assumptions/treatment.

RR means response rate, 1 means Control group and 2 means Test group. **One-tailed hypothesis** appropriate when there is interest in only one direction (Freund, 1984). For example, when taking into MPO there is a need to specify if it can increase response rate. However, **two-tailed hypothesis** is appropriate when there is a need to set the difference in any direction (Freund, 1984), for example to define if the test group will have a higher or lower response rate. In this case managers decided that both approaches can be useful. From statistical point of view there is a difference in defining the constraints of area under the curve. In the Figure 3.5.2 there are normal distributions for both hypothesis types presented.

*Figure 3.5.2 Normal distribution for one-tailed and two-tailed hypothesis.*




The third point focuses on defining the Test and Control groups, which has already been presented while defining the business objective. A control group is needed to determine effect of a marketing action. All conditions for this group should be strictly the same as for the Test group, and different from than the marketing offer. Sampling to select Test and Control groups should be random (David, 1949), but in the context of this research and the company investigated, the problem about how many customers the managers want , or rather how many customers the product managers are able to devote to participate in the testing stage is often the most decisive factor. Since these groups are perceived by the management side of the company as groups of customers who are lost in a given marketing action, this is the reason why sample size constraints and minimum sample size is so important from a business point of view. To complete this point successfully, there is a need to know when null hypothesis can be rejected or when it cannot be rejected and what type of errors it is possible to get in test outcomes.

*Table 3.5.2 Test Conclusions.*

| | Difference | No Difference |
|---|---|---|
| Reject Null hypothesis | True Positive | Type I Error |
| Do not reject Null hypothesis | Type II Error | True Negative |

If test conclusions give the result to **reject null hypothesis** it means to conclude that test group is different from the control group. If test conclusions **do not reject null hypothesis** it means to conclude that there is not enough evidence to claim that the test for group is different from the control group. Minimum sample size is counted according to the following formula (Devore, 2008):

Null hypothesis: $RR_1$-$RR_2$=0

Alternate hypothesis: $RR_1$-$RR_2$>0

$$n = \left[ z_\alpha \sqrt{\frac{(RR_1+RR_2)(Q_1+Q_2)}{2}} + z_\beta \sqrt{RR_1 Q_1 + RR_2 Q_2} \right]^2 / d^2 \qquad (3.5.1)$$

Where *n*=sample size needed, $RR_1$=Response rate in Test group, $RR_2$=Response rate in Control group, $Q_1$=(1-Response rate in Test group), $Q_2$=(1-Response rate in Control group) and *d* is the difference in Response rates that is of interest, which is significant. It is also necessary to define ***α*-probability** of concluding Response rates are different when they are actually the same, simply means **Type I Error** is probability of *α*, and to define ***β*-probability** of concluding Response rates are the same when they are actually different, simply means **Type II Error** is probability of $\beta$, and to define ***z*-distance** from mean in standard deviations. Taking into consideration $\alpha$ value it is necessary to think about the probability of falsely rejecting null hypothesis which has to be taken. Generally $\alpha$ of 0.05 value is used but if there is possibility to take more risk than $\alpha$=0.1 can be used. If only minimal risk can be taken into account then $\alpha$ of 0.01 may be reasonable**. Power** of the test is the probability that it will detect that the Control and Test groups are different, when they are actually different (by at least the effect size). It can be counted like (1-β). **Effect size** is the difference specified by the alternate hypothesis. It is worth to add that lower the probability of error is wanted, the larger the '*z*' values and larger the sample size is needed. The smaller the difference is wanted to be detected, the larger the sample size should be. To detect half of difference, the sample size should be four times larger. If there are hard constraints on how much sample size is available, it is better to look at how sample size impacts on all these parameters. In this case **the smallest minimum sample size** is the best accepted one. In the Figures 3.5.3 and 3.5.4 an illustration of the metrics above is shown. It can be also observed that sample size is 40K, effect size of interest assumes to be 0.4%. If the difference between the two groups is more than this, the test means that two groups are different. When $\alpha$=0.05 null hypothesis will be rejected at a Response rate = 2.13%. This leads to a power of 0.99 at the given effect size (area under red curve to the right of the black line). On the second below chart sample size assumes to be 10K and effect size of interest is equal 0.4%. If the difference between the two groups is more than this, it means that the two groups are different. To keep $\alpha$=0.05 null hypothesis will be rejected at a response of 2.23%. This leads to a power of 0.86 at the given effect size (as in the previous example it is area under red curve to the right of the black line).

*Figure 3.5.3 Illustration of Type I Error, Type II Error, Power and Effect Size (1).*

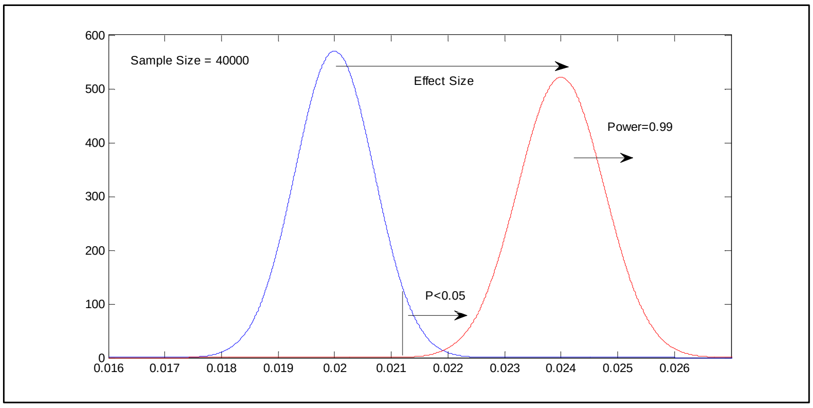
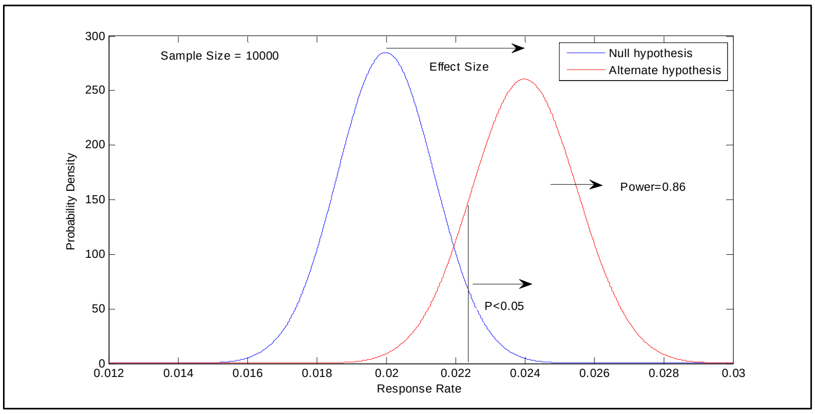



*Figure 3.5.4 Illustration of Type I Error, Type II Error, Power and Effect Size (2).*

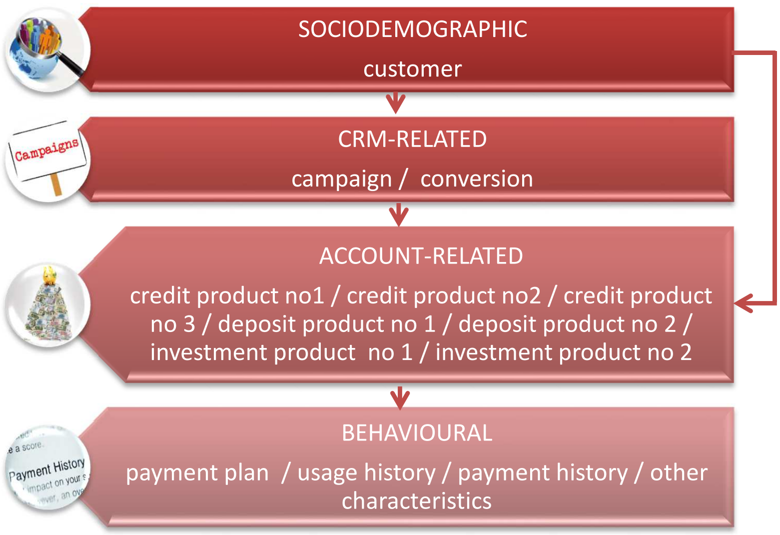


# 4. Case study - Developing the MPO campaign

## 4.1 Data overview

The data necessary to build econometric propensity-to-buy models and segmentation is stored in the form of **a relational database** on the Server in the company in the East-central Europe. A general overview of their organization is given in Figure 4.1.1.

*Figure 4.1.1 Bank database structure.*



The various sections of the Database are described below, starting from the top of the scheme:

**Customer-level data** - each customer is uniquely identified in the database using an unique RD. These RDs allow to have a single customer view and track each customer's activity through the lifetime of his relationship with the bank. The available customer-level data can include basic fields such as age, gender, location, income and marital status. For some customers, one may also have details such as the number of children, employment type and position etc. Most of the information regarding the customer is stored in one table. When a model is going to be built there is the possibility to create around seventy variables connected with customer.

**Relationship-related data** - once the customer has been acquired by the company, he may be targeted at various times with cross-sell offers. It may be that the number and type of offers made in the past, as well as the customer's response to these offers, can have a bearing on his/her future response. Therefore, variables pertaining to the relationship are used as well during modeling. For each campaign, the bank maintains a list of the customers who have been targeted, and what has been offered to them. The information on product conversions in past campaigns is also important and is stored in the table. Keys of the table there are unique customer number, unique branch number where customer has bought the product for the first time, as well the unique number of the campaign in which the customer has bought the product within. This information, when combined with the list of targets in any given campaign allows analysts to achieve some summary information on the CRM history of customer. During modeling process approximately one hundred and forty variables have been created in order to prepare one propensity-to-buy model.

**Account-level data** – every product that the customer has taken with the retail bank is recorded in the database. Each product (credit product or deposit product) can be uniquely identified using a product key. For instance, in case of credits, the product key is usually a combination of two fields – an unique number of the product and an unique number of the branch where the customer bought the product. A table containing the credit products that have been taken by all the bank's customers also contains the unique customer RD which allows analysts to link the credit products with the customers who took them. This means that it is possible to find out how many credit products were taken by the customer, what types of credit products they were, when they were taken and what the contract terms were then. Most of this information can be obtained from one table, but sometimes companies stored every product in separate table. Everything depends on the organization of the data on the server. In the propensity-to-buy model building process there is around one hundred and fifty variables created. In the propensity-to-buy model building process approximately one hundred and fifty variables have been created.

**Behavioural data** - for every product that the customer buys, there is a repayment schedule in case of credit products, and a morbidity schedule in case of a deposit portfolio. This information is often stored in a separate table. Each entry in this table is identified by the unique product key and an installment number in case of a credit product, and the number of days in case of deposit products. It contains information on how much the customer has to pay, or how much the customer paid in the beginning of saving, the split-up of this amount into principal and interest, and the payment due/maturity date. This table also tracks the actual repayment behaviour of the customer. In this context for credit products it is possible to know how much of each installment he/she paid, how it splits up into principal, interest and late payment penalty, and when he/she made the payment. As far as deposit products are concerned it is possible to know what amount of interest and when it was charged to the customer, what he/she did with it, if he/she broke up contract before the maturity date, if he/she left the money for another period of time. By matching the payment/morbidity plan with the payment/morbidity history, it is easy to know, as of any given date, what the current status of the loan/deposit is. Also by matching these two tables the delinquency status of the customer in the past can be determined. For measuring the profitability of a cross-sold loan/deposit as well, it is necessary to match the payment/morbidity plan and payment/morbidity history of the cross-sold loan/deposit and see how the customer has performed. It is also possible to define customer credit availability and customer delinquency based on payment history and delays in repayment. Such type of data is usually connected with other teams, responsible for customer risk, but it can be easily added to tables stored by CRM. If one wants to sum up the number of possible variables it will be around one hundred and fifty to be created.

As mentioned above, in order to build models and segmentation and to change the campaign approach from single-product offering to MPO and finally product-centric into customer-centric company very detailed attributes of each individual customer are needed. The solution, which is currently applied and involves storing customer attributes in several very large tables, makes gathering customer data from marketing campaigns very time consuming, since it involves collecting customer data from almost every table which is set on the server. Therefore, in view of large usage in response models building projects, there is a strong technical recommendation to define and build an **Analytical Data Mart**[5], which is to collect all data in one server space necessary to building the analytical profile and propensity-to-buy models. Obviously, its main purpose is to shorten the time which is now used to build train and validate data sets, and to decrease the number of processes, database queries and disc space.

Given the importance of the decisions that are taken by managers (e.g. interactive CRM) based on the results of data analysis, it is necessary to ensure that the results are reliable and appropriate for consideration. The

[5] A data mart is the access layer of the data warehouse environment that used to get data out to the users. The data mart is a subset of the data warehouse that it is usually oriented to a specific business line or team. Data marts are small slices of data warehouse. In this particular context Analytical Data Marta is a dedicated subset of all variables which are necessary to build the propensity-to-buy models in order to speed up the processes and to collect all available variables in one place.

observation of certain rules relating to data, analysis and organization must be ensured so that the opportunities offered by CRM analytics can be optimally utilized. In terms of data, the analysis can be performed on each data set, but if the quality of input data is questionable, then the results are not reliable. It is important that the data were characterized by the utmost quality, and were in line with reality, accurate, complete, as far as possible. The fulfillment of these conditions should be done by the operational CRM, because in this step data is collected. All employees should be aware of the importance that should be given to the quality of the data that they collect and enter into the database and systems. They should also be aware that any information can be a valuable source of knowledge. Company data is typically stored in different systems and different databases. For the analysis of analytical CRM data should be appropriately selected and prepared. Often, for this type of analyses a data warehouse is used. This kind of tool integrates data from different sources (transaction systems, accounting, operational CRM) to provide data necessary to the performance analyses. Experience showed that the better data are more atomic (not aggregated) since raw data can always be transformed and aggregated appropriately. In contrast, it is often difficult, if not impossible to convert aggregated data back to its original form.

## 4.2 Scoring

Properly conducted analysis should take into account all available data that may be relevant to the phenomena of interest. Relevant data should be analyzed using appropriate tools. Due to the nature of the issues in the bank context, the most relevant techniques in this case, however, are the most recommended by bank experts are data mining methods, which consist i.e. predictive modeling, clustering or multivariate statistical analysis.

In addition to these **advanced tools** traditional statistics are used very often as well. Drilling and exploring data tools should offer a comprehensive set of analytical techniques, which are needed for each stage of the analysis: from the data preparing, their preliminary analysis by the main part of the analysis, and the generation of final reports. It is also important that the obtained results (mainly models) can be easily **applied to new data** which were not used to develop the model. Due to the occurrence of unusual analytical problems, data mining systems are generally flexible and allow users to adjust them to the needs of the analysis.

In this research **data mining** techniques are used in the first steps of proposed empirical design. The first step **–** **scoring** - is based on the logistic regression modeling process described in chapter 3.2. This process is a kind of predictive modeling and it is defined as the process of choosing data and mathematical formulas to estimate a quantity of interest what specifically in this context means to predict propensity-to-buy score for each customer and for each product. As a result, in the customer-oriented approach the customer is described with all possible propensity-to-buy scores. So far supervised / predictive models, classified as propensity-to-buy models support targeted marketing campaigns which realize product-centric approach.

Taking into account the product managers` needs five propensity-to-buy models have been built. As these five models describe in fact four key bank products it could be summarized, that **four propensity-to-buy models** have been built:

1. Propensity-to-buy model for a **Credit Product no 1**, given that Credit Product no 1 is offered to the customer.
2. Propensity-to-buy model for a **Credit Product no 2**, given that Credit Product no 2 is offered to the customer.
3. Propensity-to-buy model for a **Credit Product no 3**, given that Credit Product no 3 is offered to the customer. This bank product is associated closely with other specific product. The relation between

these two products is based on a rule that customer needs to buy this other product before he/she decides to buy Credit Product no 3. In fact there are two possibilities (customer with or without complementary previous product), so two propensity-to-buy models, depending on the customer portfolio were built.

4. Propensity-to-buy model for a **Deposit Product no 1**, given that Deposit Product no 1 is offered to the customer.

A propensity-to-buy models building process is based on standard modeling process using logistic regression, described in the previous Chapter.

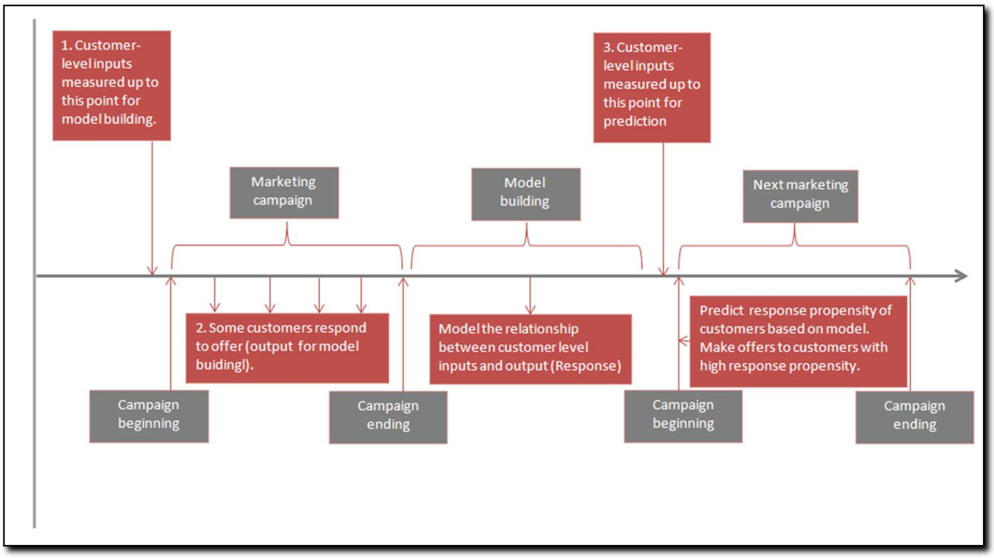
The main **goal** of each propensity to buy model presented in this study is supporting the CRM strategy for cross-selling mailing marketing campaigns in order to **increase the response rate**[6] – the main indicator which informs managers about the campaign success. By targeting only customers who are the most likely to buy given Credit or Deposit Product costs are smaller since customers who are not interested in buying are not targeted: they do not get any mailing contact from the bank.

Models are typically built on the basis of historical data from historical campaigns regarding the behavior of customers in similar situations. An example of such a process in the case of response to a product offer is presented in Figure 4.2.1. The model of customer probability to respond to a marketing offer during a marketing campaign, which means buying at least one offered product, is built by linking the customer inputs (no 1 in Figure 4.2.1) to the output, called Response (no 2 in Figure 4.2.1). Customer Response divided by number of customer in the marketing campaign gives the response rate. When the model is built, it is applied on customer-level inputs (no 3 in Figure 4.2.1) in a new campaign, and the predictions are used to pick the best customers to target in that campaign

[6] Response rate in marketing field, particularly in CRM strategy (also in the context of this research) refers to the number of people who have responded to the marketing offer = bought at least one product afer having had received the offer, divided by the number of people in the marketing campaign (number of people who were offered). It is usually expressed in the form of a percentage.

(where "the best" means the most likely, with the highest probability to buy the offered product).

*Figure 4.2.1 Propensity-to-buy modeling example scheme.*




One problem associated with propensity-to-buy modeling in retail banking is that the event which bank managers are most interested in is the event where yhe customer is buying the offered product, which does not occur often. The rate of the occurrence is called, as it is expected, response rate, and statistically – **Event Rate**. In fairly mature retail banking it is very common to note **response rates** as low as **1-5%** or less as response for marketing offers. Often, such a problem is also associated with an asymmetric payoff for correct identification of the two classes. For the instance, if **the rare event** to be modeled is that of a customer responds positively to marketing action, then the profit to be gained from the event is usually higher than the amount saved in marketing costs in case when this customer was ignored. This means that the performance metrics relevant to such a problem are more likely to focus on the ability to identify the rare responses correctly, rather than the number of non-events that are misclassified.

Logistic regression method faces one difficulty. Objective, which is pursued while finding the estimation of independent variables, is that of maximizing the likelihood function. While this is a well behaved function, it does not explain directly the business objective of the interest. For example, according to the managers, the business objective which should be achieved is to create such a data mining model, which after applied to the campaign will capture the most of all noted responders in the **top deciles**. From an analytical point of view it means to create such a data mining model where expected propensity-to-buy bank product is highly correlated with the real probability  A decile represents 10% of the database, while decile no 1 means the customers with the best scores, because to get database cut into 10 deciles customers should be **ranked** according to the estimated probability to buy (score). In practice, where it is unusual situation it is possible that the maximum likelihood estimator does not provide the best model for this goal. Therefore, as **biasing training sample** in favour of the customer response is considered, such that new biased sample has a bigger proportion of observation of events comparing to the whole training sample. The use of the biased samples for training in the **rare event problem** is quite common in practice and it allows the algorithm to model the separation between the response and the non-response. To be more precised, such approach decreases the risk that model will be over-learned or under-learned, depending on the characteristics of the biased sample. This method was used in every built model to create 10 biased samples and then to develop 10 response models. A proportion of events to non-events in every biased sample was the same 20:80. The part of events was the same in each sample since it was the whole available set of responders. The part of non-events was get by randomizing the whole set of non-responders. Then, each model was validated on the whole data set. According to finalizing model paragraph, most accurate model which met all the correct model elements requirements, was chosen.

### 4.2.1 Propensity-to-buy Credit Product no 1 model

This section focuses on the general results of created propensity-to-buy Credit Product no1 model. All the detailed information about the model is described in Appendix B.

This model was built as the first one from all models mentioned in this chapter and as it could be expected it is based on the historical data of previous actions regarding the behavior of customers in similar situation. Its goal is to predict how much a given customer is interested in buying Credit Product no 1 within the marketing action. There were many marketing campaigns which took place during the whole year before the time of building model, since campaigns had to happen regularly. As a result it was like six main campaigns with mailing contact and six followering campaigns (additional campaign to the same customers but with the other form of contact) with phone contact in the input data set. Observation window, in numbers included 1 233 454 customers who were targeted, 39 553 who filled in the credit application and 9 663 who really bought Credit Product no 1. It gives response rate equalling 3.16% for customers with a filled in application and 0.77% for real buyers. People who filled in the credit application became the target variable=1 means that Event was there noted. The rest, who was offered and did not use it become target variable=0 means that Nonevent was there noted. A model was built by using SAS 9.1 BASE tool. This statistical tool generates a lot of results. Based on these results, in Table 4.2.1.1 called ‘Response Profile’, there are listed the response categories: Event (1-‘ones’) and Nonevent (0-‘zeros’), when grouped data are input, their ordered values starting from 1 (that is why there is impossible to call nonevent like binary target variable: 0 and 1), and their total frequencies for the given data.

*Table 4.2.1.1 Response Profile for CP1.*

Response Profile

| Ordered Value | CP1 | Total Frequency |
|---|---|---|
| 1 | Event \| 1 | 39 553 |
| 2 | NonEvent \| 0 | 1 193 902 |

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied

To investigate the relationship between binary responses (target variable: 0 and 1) logistic regression was used. The logistic linear model estimated the customer willingness of purchasing Credit Product no 1 by vector of explanatory variables, after transformations has the following form:

$$P_{CP1} = Prob(Y_{CP1}|X_{CP1}) = \frac{1}{1+e^{-L_{CP1}}} \qquad (4.2.1.1)$$

where

$$L_{CP1} = \alpha + \beta_1 X_{1CP1} + \beta_2 X_{2CP1} + \cdots + \beta_k X_{kCP1} \qquad (4.2.1.2)$$

where *CP1* means Credit Product no1.

The parameters estimated of a logistic regression can be interpreted easily and in terms of **odds ratios**. The advantage of this measure of association is that it is independent of the way in which the data were collected. If the explanatory variable has more than two levels the estimated parameter can be interpreted by calculating more odds ratios. If more explanatory variables are present in a model, as in describing propensity-to-buy Credit Product no 1 model the odds ratio for one predictor may be calculated keeping all other predictors at fixed level. For a continuous explanatory variable, the odds ratio corresponds to a unit increase in the explanatory variable. Odds ratios estimates for parameters used in propensity-to-buy for Credit Product no 1 model can be foundare able to find in the below tablethe table below below (Variables ending with 'd' means that this variable is dummy variable of the original one.)

*Table 4.2.1.2 Odds ratio estimates for CP1.*

| Odds Ratio Estimates | | | |
|---|---|---|---|
| Effect | Parameter description | Point Estimate | 95% Wald Confidence Limits |
| CP1_v1_d4 | customer has got given products in rare group of products | 0.565 | 0.537 0.595 |
| CP1_v2_d4 | sum of customer debts is more than 3000 PLN | 1.133 | 1.106 1.161 |
| CP1_v3_d1 | the original interest rate of the first product | 1.244 | 1.105 1.401 |
| CP1_v3_d2 | the original interest rate of the first product is >0 and <6.3 | 1.618 | 1.432 1.828 |
| CP1_v3_d3 | the original interest rate of the first product is >6.3 and <10.3 | 1.554 | 1.373 1.759 |
| CP1_v3_d4 | the original interest rate of the first product is >10.3 | 1.811 | 1.609 2.039 |
| CP1_v4_d1 | risk grade is >0 and <0.07 | 0.459 | 0.407 0.519 |
| CP1_v4_d2 | risk grade is >0.07 and <0.13 | 0.451 | 0.403 0.505 |
| CP1_v4_d3 | risk grade is >0.13 and <0.19 | 0.481 | 0.432 0.536 |
| CP1_v4_d4 | risk grade is >0.19 and <0.25 | 0.473 | 0.425 0.526 |
| CP1_v4_d5 | risk grade is >0.25 and <0.29 | 0.451 | 0.404 0.503 |
| CP1_v4_d6 | risk grade is >0.29 and <0.38 | 0.516 | 0.465 0.571 |
| CP1_v4_d7 | risk grade is >0.38 | 0.626 | 0.568 0.690 |
| CP1_v5_d1 | customer doesn not have any credit products reported in credit bureau | 0.898 | 0.864 0.935 |
| CP1_v5_d2 | customer does not have any credit products with previous main bank | 0.821 | 0.798 0.845 |
| CP1_v5_d3 | customer does not have any credit products no 2 reported in credit bureau | 0.921 | 0.887 0.956 |
| CP1_v6_d1 | customer has got no delays in paying off the loan | 1.236 | 1.175 1.301 |
| CP1_v6_d2 | customer has got delay in paying off the loan (1-19 days) | 1.437 | 1.341 1.539 |
| CP1_v7_d1 | maximum days of delay in paying off the loan =0 | 0.779 | 0.757 0.801 |
| CP1_v7_d2 | maximum days of delay in paying off the loan is 1-2 | 0.853 | 0.821 0.887 |
| CP1_v7_d3 | maximum days of delay in paying off the loan is 3-4 | 0.908 | 0.876 0.941 |
| CP1_v8_d23 | customer deposit balance is <1500 PLN | 1.386 | 1.339 1.433 |
| CP1_v8_d36 | customer deposit balance is >10000PLN | 0.678 | 0.634 0.725 |
| CP1_v9_d43 | customer credit balance is < 1000 PLN | 1.351 | 1.286 1.419 |
| CP1_v10 | customer was communicated via phone in last year | 0.972 | 0.954 0.991 |
| CP1_v11 | customer was communicated in given season | 1.380 | 1.348 1.413 |
| CP1_v12_d1 | customer has not bought credit product no 1 in bank campaign | 0.767 | 0.730 0.806 |
| CP1_v13_d8 | customer has bought any product in bank campaign | 1.188 | 1.143 1.234 |
| CP1_v14_d1 | customer income <635 PLN | 0.764 | 0.698 0.837 |
| CP1_v14_d3 | customer income >1200 PLN | 1.064 | 1.029 1.100 |
| CP1_v15_d3 | customer is single or married | 0.820 | 0.796 0.844 |
| CP1_v16_d1 | customer graduated given education no 0 | 0.502 | 0.481 0.524 |
| CP1_v16_d2 | customer graduated given education no 1 | 1.757 | 1.707 1.808 |
| CP1_v17_d1 | any information about customer`s children | 0.441 | 0.429 0.453 |
| CP1_v17_d2 | customer has got 1 child | 1.083 | 1.052 1.115 |
| CP1_v18_d2 | customer has got given agreement type | 0.839 | 0.785 0.896 |
| CP1_v18_d5 | customer has got given agreement type | 0.703 | 0.685 0.722 |
| CP1_v19_d5 | customer lives in given district | 1.050 | 1.020 1.081 |
| CP1_v20_d1 | customer works in given occupation | 0.784 | 0.757 0.812 |
| CP1_v20_d3 | customer works in given occupation | 0.764 | 0.745 0.782 |
| CP1_v21_d1 | customer has got internet bank service access | 1.250 | 1.218 1.283 |
| CP1_v22_d1 | customer has got no dependent persons | 0.789 | 0.744 0.837 |

| | | | | |
|---|---|---|---|---|
| CP1_v23_d1 | credit risk bureau rate =0 | 0.682 | 0.664 | 0.700 |
| CP1_v23_d4 | credit risk bureau rate >3 and <5 | 1.083 | 1.048 | 1.118 |
| CP1_v24_d3 | customer lives in given district | 0.879 | 0.840 | 0.919 |
| CP1_v25_d3 | customer lives in given district | 0.535 | 0.488 | 0.586 |
| CP1_v25_d4 | customer lives in given district | 0.643 | 0.613 | 0.673 |
| CP1_v25_d5 | customer lives in given district | 0.698 | 0.667 | 0.730 |
| CP1_v25_d6 | customer lives in given district | 0.749 | 0.687 | 0.816 |
| CP1_v25_d7 | customer lives in given district | 0.794 | 0.744 | 0.848 |
| CP1_v25_d8 | customer lives in given district | 0.793 | 0.765 | 0.822 |
| CP1_v25_d9 | customer lives in given district | 0.880 | 0.853 | 0.908 |
| CP1_v25_d10 | customer lives in given district | 0.940 | 0.913 | 0.968 |
| CP1_v25_d15 | customer lives in given district | 1.136 | 1.094 | 1.178 |
| CP1_v25_d16 | customer lives in given district | 1.174 | 1.115 | 1.236 |
| CP1_v25_d17 | customer lives in given district | 1.280 | 1.231 | 1.331 |
| CP1_v26_d5 | customer has given zip code of his/her place of living | 0.893 | 0.865 | 0.921 |
| CP1_v26_d6 | customer has given zip code of his/her place of living | 0.925 | 0.899 | 0.952 |
| CP1_v26_d7 | customer has given zip code of his/her place of living | 0.866 | 0.825 | 0.909 |
| CP1_v26_d8 | customer has given zip code of his/her place of living | 0.973 | 0.948 | 0.999 |
| CP1_v26_d11 | customer has given zip code of his/her place of living | 1.140 | 1.068 | 1.216 |
| CP1_v26_d1 | customer has given zip code of his/her place of living | 1.353 | 1.324 | 1.383 |
| CP1_v27_d11 | customer is at 18-27 age | 1.184 | 1.142 | 1.227 |
| CP1_v28_d52 | customer new income >2500 PLN and <3500 PLN | 1.079 | 1.051 | 1.107 |
| CP1_v28_d53 | customer new income >3500 PLN and <4500 PLN | 1.129 | 1.087 | 1.173 |
| CP1_v28_d54 | customer new income >4500 PLN | 1.144 | 1.107 | 1.183 |
| CP1_v29_d55 | months of customer bank history <6 months | 2.456 | 2.230 | 2.706 |
| CP1_v29_d56 | months of customer bank history 6-12 months | 1.803 | 1.702 | 1.909 |
| CP1_v29_d57 | months of customer bank history 12-24 months | 1.522 | 1.461 | 1.586 |
| CP1_v29_d58 | months of customer bank history 24-36 months | 1.491 | 1.436 | 1.549 |
| CP1_v29_d59 | months of customer bank history 36-48 months | 1.266 | 1.218 | 1.315 |
| CP1_v29_d60 | months of customer bank history 48-60 months | 1.165 | 1.123 | 1.209 |
| CP1_v29_d61 | months of customer bank history 60-72 months | 1.099 | 1.060 | 1.139 |
| CP1_v29_d65 | months of customer bank history >108 months | 0.934 | 0.908 | 0.960 |
| CP1_v30_d69 | months since last bought product >24 and <36 months | 0.629 | 0.610 | 0.648 |
| CP1_v30_d70 | months since last bought product >36 and <48 months | 0.439 | 0.422 | 0.457 |
| CP1_v30_d71 | months since last bought product >48 and <60 months | 0.373 | 0.355 | 0.391 |
| CP1_v30_d72 | months since last bought product >60 and <72 months | 0.356 | 0.334 | 0.380 |
| CP1_v30_d73 | months since last bought product >72 and <84 months | 0.408 | 0.370 | 0.449 |
| CP1_v30_d74 | months since last bought product >84 months | 0.495 | 0.445 | 0.552 |
| CP1_v31 | customer has got given credit group of products no 1 | 0.608 | 0.578 | 0.640 |
| CP1_v32 | customer has got given credit product no 1 | 2.350 | 2.270 | 2.433 |
| CP1_v33 | customer has got given deposit product no 2 | 1.152 | 1.088 | 1.219 |
| CP1_v34 | customer has got given credit product no 4 | 0.836 | 0.812 | 0.862 |
| CP1_v35_d5 | customer has got given credit group of products no 2 | 0.431 | 0.412 | 0.450 |
| CP1_v36_d7 | customer has got given deposit product no 1 | 0.541 | 0.506 | 0.578 |
| CP1_v37_d28 | customer has got given credit product no 5 | 1.178 | 1.052 | 1.320 |
| CP1_v38_d32 | customer has got given credit product no 6 | 1.129 | 1.049 | 1.216 |
| CP1_v39_d33 | customer has got given deposit group of products no 1 | 1.078 | 1.013 | 1.147 |
| CP1_v40_d36 | customer has got given credit group of products no 3 | 1.305 | 1.259 | 1.353 |
| CP1_v41_d1 | sum of credit group of products no 1 =1 | 1.221 | 1.181 | 1.263 |

| | | | | |
|---|---|---|---|---|
| CP1_v41_d6 | sum of credit products no 1 is >2 but <6 | 1.123 | 1.094 | 1.152 |
| CP1_v42_d1 | sum of credit group of products no 4 =1 | 0.887 | 0.868 | 0.906 |
| CP1_v42_d3 | sum of credit group of products no 4 >3 | 1.169 | 1.122 | 1.219 |
| CP1_v43_d2 | sum of credit product no 2 =2 | 1.158 | 1.116 | 1.203 |
| CP1_v44_d2 | customer has got 2 different products | 1.271 | 1.214 | 1.331 |
| CP1_v44_d3 | customer has got 3 different products | 1.379 | 1.317 | 1.444 |
| CP1_v44_d4 | customer has got 4 different products | 1.710 | 1.630 | 1.794 |
| CP1_v44_d5 | customer has got 5-10 different products | 2.036 | 1.943 | 2.133 |
| CP1_v44_d6 | customer has got 10-42 different products | 2.225 | 2.107 | 2.350 |
| CP1_v45_d1 | customer gives salary into bank account | 1.892 | 1.832 | 1.954 |
| CP1_v46 | customer is new | 1.179 | 1.130 | 1.230 |

In general, the odds ratio shows the strength of association between a predictor and the response of interest. It can vary from 0 to infinity. If the odds ratio is one, there is no association. In the propensity-to-buy Credit Product no 1 model there are seven variables for which odds ratio is just a little bit higher than 1. The remaining variables are higher than 1 (43 variables) or smaller than 1 (52 variables). For variable CP1_v33, where odds ratio=1.152 it can be interpreted as follows: if there are customers who are different from each other only in terms of portfolio and of having deposit product no 2, the customer, who had bought deposit product no 2 has an odds ratio of buying CP1 versus not buying CP1 which is 15,2% higher than for customers who did not buy deposit product no 2. For variable CP1_v30_d72, where odds ratio=0.356 it can be in turn described in a way: if there are customers who are different from each other only in terms of time since last bought product, the customer who had bought last product between 60 and 72 months ago has and odds ratio of buying CP1 versus not buying CP1 which is 64,4% lower than for customers who had bought his/her last product like benchmark category shows (earlier than 24 months ago).

Sufficient replication within subpopulation is required to make the **Pearson and deviance goodness-of-fit tests** valid. When there are one or more continuous predictors in the model, the data are often too sparse to use these statistics. That is why **Hosmer and Lemeshow** proposed a statistic that they show, through simulation, is distributed as chi-square when there is no replication in any of the subpopulations. The figures in Table 4.2.1.3 show the Hosmer and Lemeshow test results of propensity-to-buy Credit Product no1 model, which examined model fit to the observed values. All observations were divided into 10 groups according to the increasing probability, which defines the distribution of values compatible with the observed distribution of the theoretical value. The null hypothesis implies a good fitness to the data model against the alternative one meaning a bad match. In this model there is no reason to reject the null hypothesis, which allows to conclude that developed model is well suited to the data. Created 10 groups are the synonyms of deciles and their results are also presented in the table below.

*Table 4.2.1.3 The Hosmer-Lemeshow goodness-of-fit results of propensity-to-buy CP1 model.*

| Hosmer and Lemeshow Goodness-of-Fit Test | | |
|---|---|---|
| Chi-square | DF | Pr>Chi-square |
| 5.8880 | 8 | 0.6598 |

| Hosmer-Lemeshow test results | | | | | |
|---|---|---|---|---|---|
| Group | Total | Target_variable=1 | | Target_variable=0 | |
| | | Observed | Predicted | Observed | Predicted |
| 1 | 122 374 | 554 | 555.08 | 121 820 | 124 911.08 |
| 2 | 122 420 | 715 | 714.76 | 121 705 | 122 191.27 |
| 3 | 122 482 | 932 | 935.72 | 121 550 | 121 134.65 |
| 4 | 122 547 | 1160 | 1 159.60 | 121 387 | 124 113.70 |
| 5 | 122 591 | 1314 | 1 309.51 | 121 277 | 118 165.84 |
| 6 | 122 704 | 1709 | 1 715.96 | 120 995 | 120 876.96 |
| 7 | 122 850 | 2185 | 2 182.87 | 120 665 | 121 156.70 |
| 8 | 123 121 | 3 169 | 3 087.70 | 119 952 | 120 185.39 |
| 9 | 124 016 | 6 299 | 6 440.49 | 117 717 | 118 187.33 |
| 10 | 128 363 | 21 516 | 22 061.95 | 106 847 | 106 810.57 |

The **Kolmogorov Smirnov** test (KS statistics) is one of the statistics which defines the absolute difference between cumulative percentage of events and cumulative percentage of nonevents and it simply summarizes up to which decile the events have the superiority over nonevents.

$$KS_k = \sum_{i=1}^{k} f_{events} - \sum_{i=1}^{k} f_{nonevents} \qquad (4.2.1.3)$$

Above formula calculates KS statistics for *k* th decile with taking into account the frequency (*f*) of events and nonevents. In the Table 4.3.1.6 there are KS statistics for CP1 propensity-to-buy model results.

*Table 4.2.1.4 The Kolmogorov Smirnov results of propensity-to-buy CP 1 model.*

| DECILE | KS statistics |
|---|---|
| ~10% | 71,0% |
| ~20% | **80,9%** |
| ~30% | 78,2% |
| ~40% | 71,6% |
| ~50% | 63,2% |
| ~60% | 53,1% |
| ~70% | 42,4% |
| ~80% | 30,7% |
| ~90% | 18,3% |
| ~100% | 5,1% |
| maximum | 80,9% |

While looking at the table above it is easy to observe that the maximum value of KS statistics is present in 2nd decile. So in the situation of necessity to optimize the customer offering it is recommended to stop the customer solution just after the second decile.

**ROC** (Receiver Operating Characteristic) curves are used to evaluate and compare the performance of diagnostic tests. They can also be used to evaluate model fit. A ROC curve is just a plot of the proportion of the true positives (events predicted to be events) versus the proportion of false positives (nonevents predicted to be events). The bigger space between ROC line and baseline random line, the better model is going to detect Events. On the below Figure there is ROC curve for Development sample of propensity-to-buy Credit Product no1 model. It occurs that created model fits the observations and it is going to predict marketing offer responders correctly.

*Figure 4.2.1.1 ROC charts for development data sets of propensity-to-buy CP 1 model.*

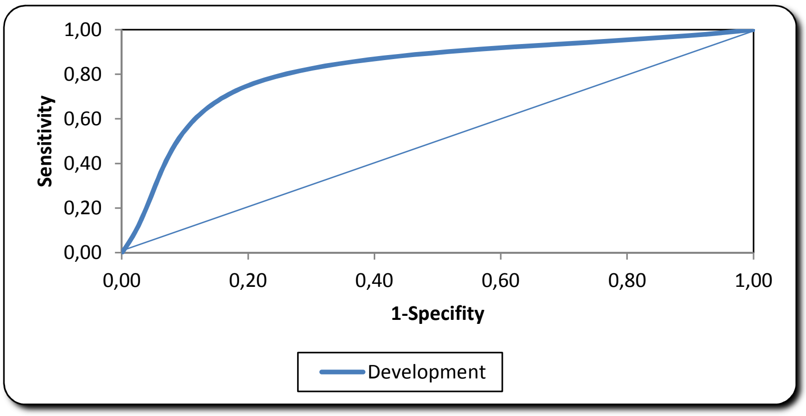


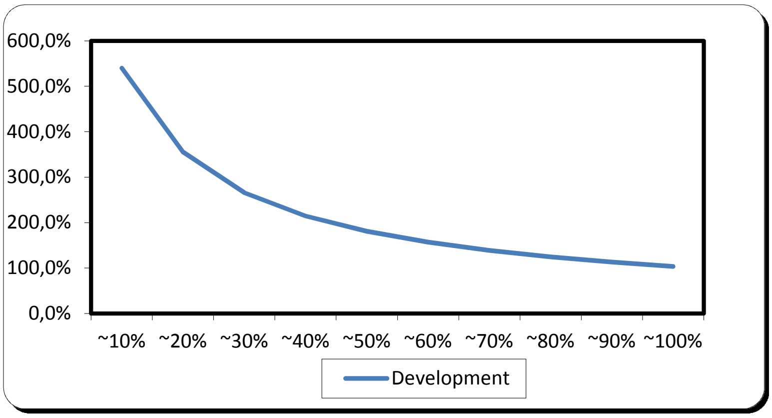

Table 4.2.1.5 shows the number of Events and Nonevents and **Lift rate** value for each decile of Development sample.

*Table 4.2.1.5 Events and Nonevents results and Lift rate for application propensity-to-buy CP1 model on development sample.*

| Development sample | | | | | | |
|---|---|---|---|---|---|---|
| Decile | Non Resp. | Resp. | Total | Resp. Rate | Cum. Resp rate | % Lift |
| ~10% | 106 847 | 21 516 | 128 362 | 16,76% | 16,76% | 540,0% |
| ~20% | 117 717 | 6 299 | 124 015 | 5,08% | 11,02% | 355,1% |
| ~30% | 119 952 | 3 169 | 123 121 | 2,57% | 8,25% | 265,8% |
| ~40% | 120 655 | 2 185 | 122 839 | 1,78% | 6,66% | 214,4% |
| ~50% | 120 995 | 1 709 | 122 704 | 1,39% | 5,62% | 180,9% |
| ~60% | 121 277 | 1 314 | 122 591 | 1,07% | 4,87% | 156,8% |
| ~70% | 121 387 | 1 160 | 122 546 | 0,95% | 4,31% | 138,9% |
| ~80% | 121 550 | 932 | 122 482 | 0,76% | 3,87% | 124,8% |
| ~90% | 121 705 | 715 | 122 420 | 0,58% | 3,51% | 113,1% |
| ~100% | 121 820 | 554 | 122 374 | 0,45% | 3,21% | 103,3% |
| | 1 193 902 | 39 553 | 1 233 454 | 3,21% | | |

It is easy to observe that decile no 1 has the biggest number of Events which implies the highest value of response rate. **Cumulative response rate** means that when deciding to cut data set in given decile, this response rate is going to be obtained. The last column shows th Lift rate, which is the most commonly used metric to measure the performance of targeting models in marketing applications. The purpose of a simple propensity-to-buy modeling is to identify a decile or deciles, or just some subgroups (target) from a larger population to make the targeting marketing sense. The target members selected are those likely to responds to a marketing product offer. Model, is doing well if the response within the target is much higher than average for the population as a whole. Lift rate is the ratio of these values: target response divided by average response. Normally, decile contains 1/10 of the whole, used to modeling population, the highest responders are put into decile 1 and the lowest into 10 decile. In the top decile of Development sample there were 128 362 customers, where were 21 516 responders with a response rate of 16.76%. Compared to the average response rate of 3.21%, this gives a lift 5.40 (540%) for decile 1. Each successive decile has a lower response rate, which is correct since it means that model is ordering customers in a proper way.

*Figure 4.2.1.2 Lift ratio results charts for development data set for CP1.*



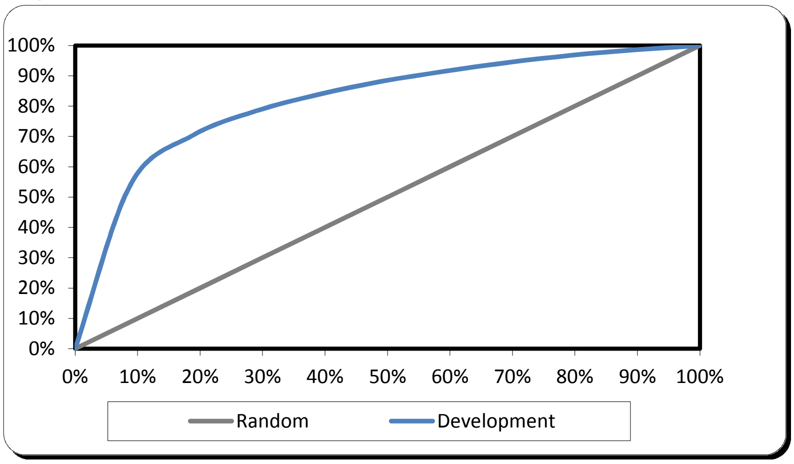

Another useful Figure above compares the cumulative percent of responses captured as each decile is added to the target.

According to the table below, the top two deciles capture 70.3% or the responders (Events). This is compared to a random baseline where two deciles (20% of the population) would capture 20% of the responders. This result is almost like the '80/20' rule, which means that it is much better than not targeting. The greater the area between two lines – baseline and line with the information of cumulative percent of Responses Captured, the more the model is able to concentrate responders in the top deciles.

*Table 4.2.1.6 Event and Nonevent distribution in deciles for development sample of propensity-to-buy for CP1 model.*

| Model Development all APPL | | | | | | | |
|---|---|---|---|---|---|---|---|
| Approx. score % | NonResp (NR) | Responders (R) | Prob. NonResp | % of all NonResp | % of all Resp | Cum. % NonResp | Cum. % Resp |
| ~10% | 106 847 | 21 516 | 0,832 | 8,9% | 54,4% | 8,9% | 54,4% |
| ~20% | 117 717 | 6 299 | 0,949 | 9,9% | 15,9% | 18,8% | 70,3% |
| ~30% | 119 952 | 3 169 | 0,974 | 10,0% | 8,0% | 28,9% | 78,3% |
| ~40% | 120 655 | 2 185 | 0,982 | 10,1% | 5,5% | 39,0% | 83,9% |
| ~50% | 120 995 | 1 709 | 0,986 | 10,1% | 4,3% | 49,1% | 88,2% |
| ~60% | 121 277 | 1 314 | 0,989 | 10,2% | 3,3% | 59,3% | 91,5% |
| ~70% | 121 387 | 1 160 | 0,991 | 10,2% | 2,9% | 69,4% | 94,4% |
| ~80% | 121 550 | 932 | 0,992 | 10,2% | 2,4% | 79,6% | 96,8% |
| ~90% | 121 705 | 715 | 0,994 | 10,2% | 1,8% | 89,8% | 98,6% |
| ~100% | 121 820 | 554 | 0,995 | 10,2% | 1,4% | 100,0% | 100,0% |
| Totals | 1 193 902 | 39 553 | | | | | |

*Figure 4.2.1.3 Comparison of cumulative percent of Events and Nonevents for development and validation data set.*



The model represents enough quality of goodness-of-fit of the set of explanatory variables to the target variable. The curve is enough above the random line.

### 4.2.2 Models summary

Since the remaining propensity-to-buy models were built according to the same method and algorithms and all building stages are the same, the descriptions of the results are presented in the Appendices.

The last step of the modeling process is Model validation. The Development set is used for learning, that is for fitting the model parameters And Validation set is used to tune these parameters and to minimize overfitting the model, which means that model correctness should be checked on the Validation data set (eg.on the data set presents data and results from the next marketing campaign).

In the Tables and Figures below **summaries of each propensity-to-buy model** are presented based on the Development and Validation data set. Summaries are expressed in the form of selected metrics.

*Table 4.2.2.1 Basic numbers of each propensity-to-buy model and development and validation data set.*

| Model | Event | NonEvent | Respone Rate |
|---|---|---|---|
| MPTB CP1 Development | 39 553 | 1 193 902 | 3,2% |
| MPTB CP1 Validation | 25 496 | 795 934 | 3,1% |
| MPTB CP2 Development | 8 109 | 1 994 342 | 0,4% |
| MPTB CP2 Validation | 4 055 | 997 171 | 0,4% |
| MPTB CP3A Development | 9 743 | 424 224 | 2,2% |
| MPTB CP3A Validation | 1 607 | 328 438 | 0,5% |
| MPTB CP3B Development | 1 431 | 853 990 | 0,2% |
| MPTB CP3B Validation | 716 | 768 591 | 0,1% |
| MPTB DP1 Development | 14 403 | 57 987 | 19,9% |
| MPTB DP1 Validation | 3 070 | 14 407 | 17,6% |

And KS statistics for all data sets for all propensity-to-buy models.

*Table 4.2.2.2 KS statistic for each propensity-to-buy model and development and validation data sets.*

| DECILE | MPTB CP1 Dev | MPTB CP1 Val | MPTB CP2 Dev | MPTB CP2 Vall | MPTB CP3A Dev | MPTB CP3A Val | MPTB CP3B Dev | MPTB CP3B Val | MPTB DP1 Dev | MPTB DP1 Val |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 71% | 50% | 62% | 59% | 29% | 35% | 41% | 38% | 49% | 52% |
| 2 | 81% | 46% | 66% | 62% | 48% | 38% | 60% | 53% | 78% | 83% |
| 3 | 78% | 37% | 62% | 57% | 55% | 39% | 61% | 51% | 76% | 82% |
| 4 | 72% | 29% | 55% | 51% | 50% | 39% | 56% | 47% | 68% | 72% |
| 5 | 63% | 21% | 47% | 43% | 44% | 34% | 48% | 41% | 57% | 61% |
| 6 | 53% | 13% | 38% | 36% | 36% | 22% | 39% | 34% | 47% | 49% |
| 7 | 42% | 7% | 29% | 27% | 27% | 13% | 29% | 26% | 35% | 37% |
| 8 | 31% | 3% | 20% | 18% | 18% | 7% | 20% | 18% | 24% | 24% |
| 9 | 18% | 1% | 10% | 9% | 10% | 4% | 10% | 9% | 12% | 12% |
| 10 | 5% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| max | 81% | 50% | 66% | 62% | 55% | 39% | 61% | 53% | 78% | 83% |

*Table 4.2.2.3 Lift rate for each propensity-to-buy model and development and validation data sets.*

| DECILE | MPTB CP1 Dev | MPTB CP1 Val | MPTB CP2 Dev | MPTB CP2 Va | MPTB CP3A Dev | MPTB CP3A Val | MPTB CP3B Dev | MPTB CP3B Val | MPTB DP1 Dev | MPTB DP1 Val |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 5,40 | 3,83 | 7,29 | 6,86 | 3,81 | 4,46 | 5,15 | 4,86 | 4,95 | 5,65 |
| 2 | 3,55 | 2,23 | 4,31 | 4,10 | 3,36 | 3,42 | 4,02 | 3,67 | 4,14 | 4,97 |
| 3 | 2,66 | 1,70 | 3,06 | 2,87 | 2,79 | 2,69 | 3,05 | 2,71 | 3,02 | 3,44 |
| 4 | 2,14 | 1,44 | 2,38 | 2,26 | 2,23 | 1,98 | 2,40 | 2,18 | 2,35 | 2,57 |
| 5 | 1,81 | 1,27 | 1,94 | 1,86 | 1,85 | 1,64 | 1,96 | 1,82 | 1,92 | 2,04 |
| 6 | 1,57 | 1,14 | 1,64 | 1,59 | 1,58 | 1,31 | 1,65 | 1,57 | 1,62 | 1,69 |
| 7 | 1,39 | 1,07 | 1,42 | 1,39 | 1,37 | 1,15 | 1,42 | 1,37 | 1,40 | 1,44 |
| 8 | 1,25 | 1,03 | 1,25 | 1,23 | 1,22 | 1,07 | 1,25 | 1,23 | 1,24 | 1,25 |
| 9 | 1,13 | 1,01 | 1,11 | 1,10 | 1,10 | 1,04 | 1,11 | 1,10 | 1,11 | 1,11 |
| 10 | 1,03 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 |

*Figure 4.2.2.1 Lift ratio results for each propensity-to-buy model and development and validation data sets.*

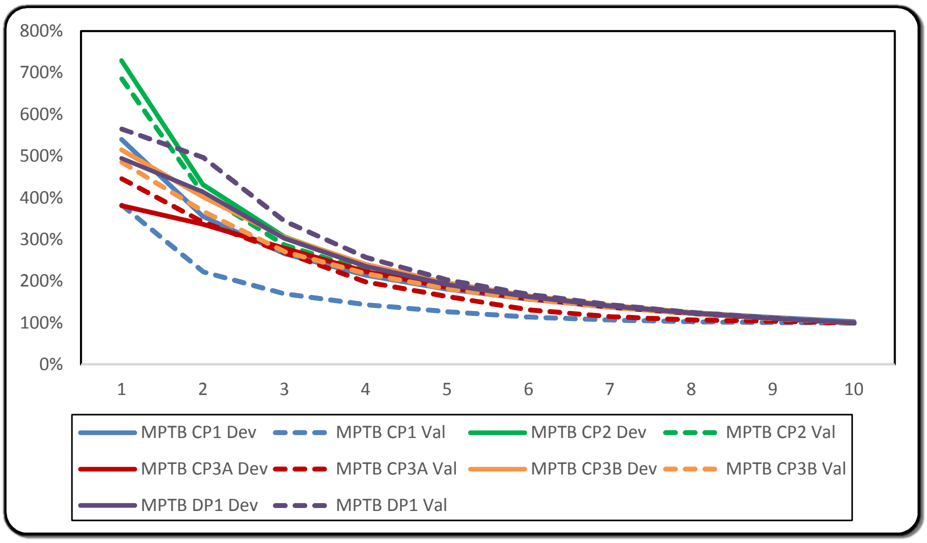
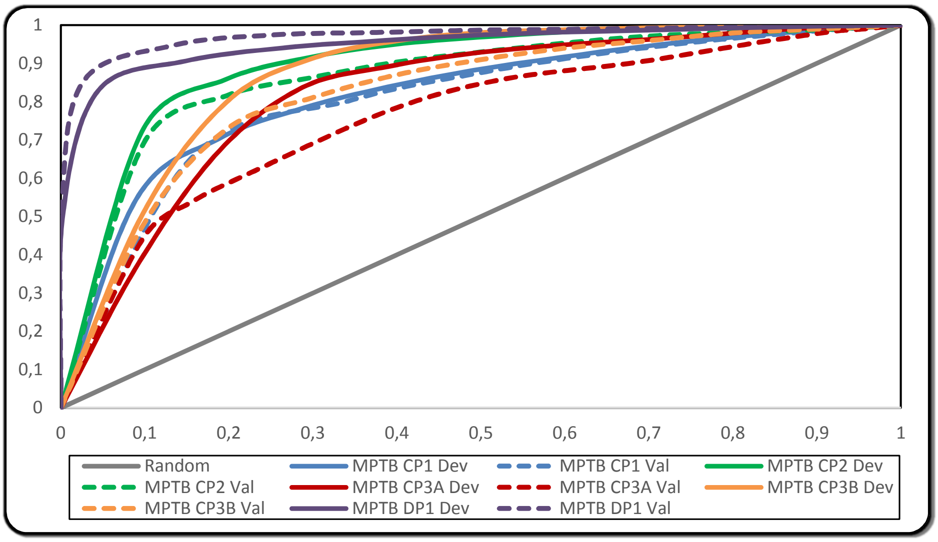



*Figure 4.2.2.2 Comparison of cumulative percent of Events and Nonevents for development and validation data set of each propensity-to-buy model.*



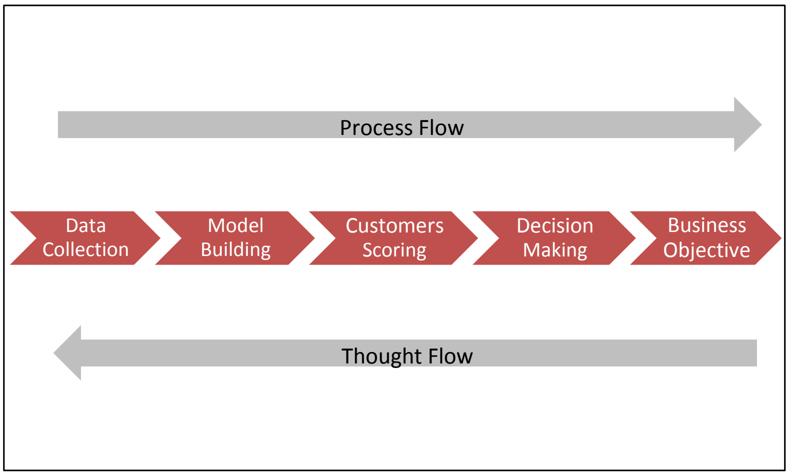

Looking at the charts and having analyzed the numbers in the tables above it could be concluded that developed propensity-to-buy models are characterized with high quality and high goodness of fit the model to the reality.

## 4.3 Segmentation and MPO

This part of the study presents the method which was used to found the solution of changing the marketing offering from single-product campaign into multiproduct campaign in order to change the organization from product-centric into customer-centric. It provides the details of the segmentation idea, which uses inputs from the previous sections as well as from the business. There are in fact two types of **inputs** required for this process. The first set of inputs are the response/buying probabilities of each particular propensity-to-buy model that has been built. The second set of inputs are obtained from the product managers. It includes a definition of the business` objective of the multiproduct offering approach. The **output** of this process is the decision on which set of products should be offered to which customer.

Earlier chapters in this document have discussed how one could get the right data to build propensity-to-buy models on them. The objective of the model is to predict how a customer will behave in the future and use these predictions to select customers to send offers to. Through these offers (or rather through the customers` response to offers), the bank achieves some desired objectives. **The decision making process** is towards the end of the entire procedure. But this is a step which determines what needs to be done in other steps because this is the closest to understanding the business needs. **The thought process** works in the opposite direction of the process flow. The thought process starts from understanding the business needs and formulation the decision-making process to fulfill the business requirements. Then, to make the decision analysis what necessary information is needed. Sometimes such information is available from the business, otherwise it has to be estimated by using predictive models. Therefore, the decision making process determined which predictive models to build and in turn determined what data are needed to build those models. In the Figure 4.3.1 this process is presented graphically.

*Figure 4.3.1 Process flow.*



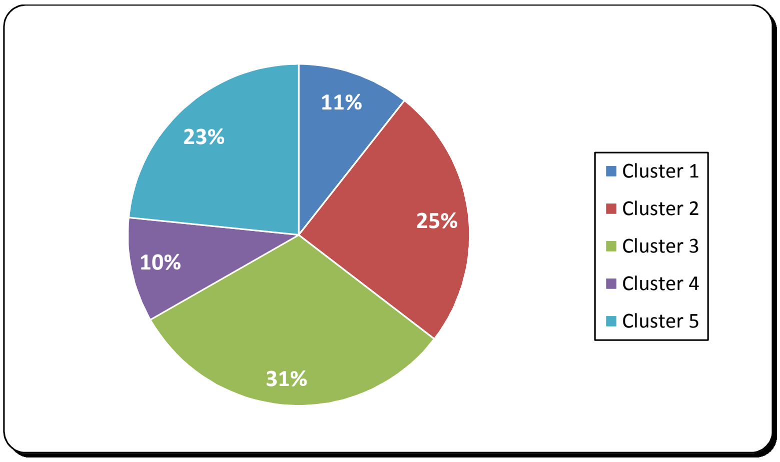

The cross-sell problem of a bank with a set of customers on its books, whom it would like to target with more than one product at the same time was considered. It was initially assumed that starting from changing the offering rules will lead to change the product-centric orientation into customer-centric. To meet business requirements and solve the problem of offering the same customers in the same order and to expand the base of the targeting customers, **customer propensity-to-buy segmentation** was built. Since the goal of this treatment was to **find similar group of customers who are likely to buy the similar portfolio** the segmentation was based on the deciles of each model and it was the input, as mentioned above. As the output this method gave five clusters of customers which were detected according to perfect portfolio (perfect at a given time) which should be offered to customers. The choice of product offer to a given customer can be said to depend on four factors:

1. The propensity-to-buy of the customer to a Credit Product no 1, expressed in ten model deciles.
2. The propensity-to-buy of the customer to a Credit Product no 2, expressed in ten model deciles.
3. The propensity-to-buy of the customer to a Credit Product no 3, expressed in ten model deciles.
4. The propensity-to-buy of the customer to a Deposit Product no 1, expressed in ten model deciles.

As discussed in the previous section, the key parameters (propensity-to-buy scores) are predicted based on the past information available about customers. As a result each key parameter gets the estimated value. It is very difficult to get accurate estimated of these parameters at the customer level. However, it is easy to observe if customers are grouped into quantiles on any given parameter. The quantile-level aggregate estimated perform better in terms of the averaged and the rank ordering. Since the quality of the solution given by the segmentation is dependent upon the quality of inputs, it is more meaningful to formulate the problem at a group level rather than a customer level.

Since in k-means method it is the user, who decides how many different clusters segmentation are going to become, there are some possible methods to support the expert. In this research a number of clusters were estimated using the method of CCC - Cubic Clustering Criterion (SAS Institute Inc., Cary, NC, USA, 1983, "SAS Technical Repot A-108, Cubic Clustering Criterion") and deep analysis of V-fold cross evaluation – repeating the procedure for drawing a sample of data to analyze and build V-times (here: 5 times). In practice, very often it is quite difficult to determine the number of clusters that is the most appropriate. Therefore, if the aim is to draw up a typology of empirical objects, looking for the appopriate number of clusters should be guided by the practical principle, and the number of clusters should be large enough to allow to extract different types of observations. It is also important to obtain different clusters from the business point of view, in the ideal situation each cluster should represent different, clear business interpretation. As a result of such activities five segments were achieved. Each segment was characterized according to the differentiating variables, i.e. the deciles of each propensity-to-buy model. The following Figure presents the size of created segments.

*Figure 4.3.2 Size of segments.*




The table below shows basic measures of every variable broken down by segment and size of segments.

*Table 4.3.1 Characteristics of segments.*

| NAME | CLUSTER | NUMBER | VARIABLE | MEAN | MEDIAN | 25th PERCENTILE | 75th PERCENTILE | RESPONSE RATE | RESPONSE RATE [%] |
|---|---|---|---|---|---|---|---|---|---|
| CP1 & CP2 | 1 | 268 396 | DECILE_CP1 | 2,15 | 2 | 1 | 3 | 10 454 | 3,92% |
| | | | DECILE_CP2 | 3,74 | 3 | 2 | 6 | 4 635 | 1,71% |
| | | | DECILE_CP3 | 7,21 | 7 | 6 | 8 | 1 008 | 0,36% |
| | | | DECILE_DP1 | 8,53 | 9 | 8 | 10 | 3 263 | 3,33% |
| CP1 & CP2 & DP1 | 2 | 628 038 | DECILE_CP1 | 2,36 | 2 | 1 | 3 | 30 380 | 4,82% |
| | | | DECILE_CP2 | 3,82 | 3 | 2 | 6 | 10 463 | 1,67% |
| | | | DECILE_CP3 | 4,44 | 4 | 3 | 6 | 17 370 | 2,75% |
| | | | DECILE_DP1 | 2,37 | 3 | 2 | 4 | 18 486 | 4,01% |
| DP1 | 3 | 792 026 | DECILE_CP1 | 6,93 | 7 | 6 | 8 | 3 326 | 0,41% |
| | | | DECILE_CP2 | 8,16 | 9 | 7 | 9 | 2 439 | 0,32% |
| | | | DECILE_CP3 | 5,73 | 6 | 4 | 9 | 10 877 | 1,40% |
| | | | DECILE_DP1 | 1,00 | 1 | 1 | 1 | 19 206 | 2,52% |
| CP1 & CP2 & CP3 | 4 | 250 331 | DECILE_CP1 | 2,28 | 2 | 1 | 3 | 13 536 | 5,40% |
| | | | DECILE_CP2 | 2,76 | 2 | 1 | 4 | 5 751 | 2,30% |
| | | | DECILE_CP3 | 2,82 | 3 | 2 | 4 | 12 303 | 4,91% |
| | | | DECILE_DP1 | 7,81 | 8 | 7 | 9 | 4 725 | 2,48% |
| NO OFFER | 5 | 591 507 | DECILE_CP1 | 7,67 | 8 | 6 | 9 | 2 930 | 0,50% |
| | | | DECILE_CP2 | 7,32 | 7 | 6 | 8 | 1 085 | 0,18% |
| | | | DECILE_CP3 | 5,28 | 5 | 4 | 7 | 1 224 | 0,23% |
| | | | DECILE_DP1 | 8,26 | 9 | 7 | 10 | 891 | 0,63% |

The table presents size of clusters and value of mean and median measure for each of differentiating variable in every cluster. It also includes value of 25$^{th}$ and 75$^{th}$ percentiles. Based on the data which was used to create five segments there were also response rates for each bank product counted and they were presented in percentage form as well. First, mean and median measures were analyzed. It was the analytical input. The smaller decile, the propensity-to-buy given product by customer is higher. Then, response rates values were carefully compared to each other and possibilities to offer two or more products in one letter were analyzed. It was the business input. The combination of these two factors (analytical and business) gave the final output which results in the names of the segments which are placed in the first column of Table 4.3.1. As it can be observed, first segment is the two-product segment, the customers are offered with two credit products – Credit Product no 1 and Credit Product no 2. Segment no 2 is three-product segment, the customers are offered with two credit products (Credit Product no 1 and Credit Product no 2) and one deposit product (Deposit Product no 1).

The third segment focused on deposit product. The customers were only offered with Deposit Product no 1. The next segment focused on the credit side of products and the customers were only offered by credit products – Credit Product no1, no 2 and no 3. The fifth and last segment included customers who were not likely to buy bank products. This group needs more investigation by product managers because these customers may need a special portfolio, other communication channel or maybe anti-churn treatment policy. Although three segments (Segment no 1, Segment no 2, Segment no 4) include customers with high propensity-to-buy Credit Product no 1, when analyzing the response rates it is possible to assume that the volume of response rate which is available to achieve is the highest in Segment no 4, followed by Segment no 2 and Segment no 1.

# 5. Case Study – Experimental Evaluation

## 5.1 Experimental Setup

The newly proposed multiproduct offering approach has to be tested before being launched as an ordinary, daily marketing cross-selling process. This section discusses how to prepare the real case study in order to check the assumptions, hypothesis and finally – to find the answers for the research questions. This subchapter presents sample size determining method. Moreover, it shows the whole test design with precise treatments provided for each distinguished group of the customers.

In the multiproduct offering (MPO) approach **minimum sample size** (Barnett, 2002) was counted according to the following steps :

1. Choose range of response rate.
2. Determine sample size constraints.
3. Select α level.
4. Select Power levels (depends on β value).
5. Calculate sample size for various effect sizes at each combination of α and Power.
6. Plot sample size vs. Power for various Effect sizes, examine tradeoffs between the various parameters to finalize sample size.

In Appendix C dependences between Power and Sample Size for α=0.05 and α=0.1 and for several values of Effect size (0.40-1.00%) are presented, in the form of several charts. They are connected with all five Segments.
After an in-depth analysis sizes of Champion and Challenger groups were specified. To facilitate the analysis of results Champion groups of each segment were decided to have the same sizes and they consist of **18 900 customers**. The same rule was applied to Control groups. Each Control group for every segment consists of **6 500 customers**. In the table below all types of Test and Control groups with their sizes and with the prepared type of marketing offer are presented.

*Table 5.1.1 Summary of created Test and Control groups.*

| NAME | TEST GROUP | CONTROL GROUP | SIZE | OFFER TYPE |
|---|---|---|---|---|
| SEGMENT_1_CL_1 | 0 | 1 | 6 500 | No offer |
| SEGMENT_1_CL_2 | 0 | 1 | 6 500 | CP2 |
| SEGMENT_1_CL_3 | 0 | 1 | 6 500 | DP1 |
| SEGMENT_1_CL_4 | 0 | 1 | 6 500 | CP3 |
| SEGMENT_1_CL_5 | 0 | 1 | 6 500 | CP1\|CP2\|CP3 |
| SEGMENT_1_CL_6 | 0 | 1 | 6 500 | CP1\|CP2\|DP1 |
| SEGMENT_1_CL_7 | 0 | 1 | 6 500 | CP1 |
| SEGMENT_1_CM | 1 | 0 | 18 900 | CP1\|CP2 |
| SEGMENT_2_CL_1 | 0 | 1 | 6 500 | CP1 |
| SEGMENT_2_CL_2 | 0 | 1 | 6 500 | CP2 |
| SEGMENT_2_CL_3 | 0 | 1 | 6 500 | DP1 |
| SEGMENT_2_CL_4 | 0 | 1 | 6 500 | CP3 |
| SEGMENT_2_CL_5 | 0 | 1 | 6 500 | CP1\|CP2\|CP3 |
| SEGMENT_2_CL_6 | 0 | 1 | 6 500 | No offer |
| SEGMENT_2_CL_7 | 0 | 1 | 6 500 | CP1\|CP2 |
| SEGMENT_2_CM | 1 | 0 | 18 900 | CP1\|CP2\|DP1 |
| SEGMENT_3_CL_1 | 0 | 1 | 6 500 | CP1 |
| SEGMENT_3_CL_2 | 0 | 1 | 6 500 | CP2 |
| SEGMENT_3_CL_3 | 0 | 1 | 6 500 | No offer |
| SEGMENT_3_CL_4 | 0 | 1 | 6 500 | CP3 |
| SEGMENT_3_CL_5 | 0 | 1 | 6 500 | CP1\|CP2\|CP3 |
| SEGMENT_3_CL_6 | 0 | 1 | 6 500 | CP1\|CP2\|DP1 |
| SEGMENT_3_CL_7 | 0 | 1 | 6 500 | DP1\|CP3 |
| SEGMENT_3_CL_8 | 0 | 1 | 6 500 | CP1\|CP2 |
| SEGMENT_3_CM | 1 | 0 | 18 900 | DP1 |
| SEGMENT_4_CL_1 | 0 | 1 | 6 500 | CP1 |
| SEGMENT_4_CL_2 | 0 | 1 | 6 500 | CP2 |
| SEGMENT_4_CL_3 | 0 | 1 | 6 500 | DP1 |
| SEGMENT_4_CL_4 | 0 | 1 | 6 500 | CP3 |
| SEGMENT_4_CL_5 | 0 | 1 | 6 500 | No offer |
| SEGMENT_4_CL_6 | 0 | 1 | 6 500 | CP1\|CP2\|DP1 |

| | | | | |
|---|---|---|---|---|
| SEGMENT_4_CL_7 | 0 | 1 | 6 500 | CP1\|CP2 |
| SEGMENT_4_CM | 1 | 0 | 18 900 | CP1\|CP2\|CP3 |
| SEGMENT_5_CL_1 | 0 | 1 | 6 500 | CP1 |
| SEGMENT_5_CL_2 | 0 | 1 | 6 500 | CP2 |
| SEGMENT_5_CL_3 | 0 | 1 | 6 500 | DP1 |
| SEGMENT_5_CL_4 | 0 | 1 | 6 500 | CP3 |
| SEGMENT_5_CL_5 | 0 | 1 | 6 500 | CP1\|CP2\|CP3 |
| SEGMENT_5_CL_6 | 0 | 1 | 6 500 | CP1\|CP2\|DP1 |
| SEGMENT_5_CL_7 | 0 | 1 | 6 500 | DP1\|CP3 |
| SEGMENT_5_CL_8 | 0 | 1 | 6 500 | CP1\|CP2 |
| SEGMENT_5_CM | 1 | 0 | 18 900 | No offer |
| TOTAL | 5 | 37 | 335 000 | |

In total, **335K customers** participated in the multiproduct offering approach test, which was based on proposed propensity-to-buy scores expressed in deciles segmentation. Customers were offered all possible combinations to check if proposed approach was working correctly and if MPO was able to achieve higher response rates than single-product offering, which is characteristic for the product-centric organization.

## 5.2 Research questions

Most importantlyof all, the goal of this reseach is to change product-centric organization into customer-centric by changing marketing customer treatment as athe first step. It is realized by changing the marketing customer offering from single-product offering into MPO. The MPO test has been designed and created as ain the way to find the answer tofor the main research question as tobout which approach can achieve better results – the existing single-product offering approach or the newly proposed MPO approach. This comparison is expressed by comaparing response rates for each bank product offered in the campaign. All theseis efforts aimleads to **meet product and cross-sell managers expectations** because they are those individuals who are the most interested in the results achieved and the recommendations stemming from the conducted MPO test. In the context of the bank where the research was taken, the Product Manager is one who takes care of the one bank product. He/she is responsible for launching bank product in the market and he/she is responsible for the highest sales results of this product as well. But above all, he/she is the person who decides how many customers should be offered this product to retain the estimated sales plan and to achieve the best results. That is the reason why he/she often relies on the propensity-to-buy models and the scores which are estimated by the models. The easiest way to satisfy them is to present the obtained results in the form of the answers to the **research questions** that had been defined.

MPO test results can be described from the several perspectives for each of four offered products: from the whole marketing campaign perspective (1), from the created segments perspectives (2), from the type of the communication addressed to the customer: multiproduct with given product, given single-product, no offer, any other offer without given product (3), from the particular, the exact kind of multiproduct combination with given product (CP1|CP2, CP1|CP2|CP3, CP1|CP2|DP1) (4). By combining these four perspectives four main Research Question dedicated to the achieved Results are formulated:

**Research Question no 1 (based on the (1)(3)):** How high response rates are obtained in the whole campaign (MPO test) for each of four proposed bank products? Is there any difference between response rates achieved in the groups of customers with different types of communication – single-product offering, multiproduct offering with given product, any other communication without given product, no communication at all?

**Research Question no 2 (based on the (1)(4)):** Which of the created combination of products in multiproduct offer represents the higher response rate from the given product perspective in the whole campaign?

**Research Question no 3 (based on the (2)(3)):** How high response rates are obtained in the particular segments for each of four proposed bank products? Is there any difference between response rates achieved in the groups of customers with different type of communication – single-product offering, multiproduct offering with given product, any other communication without given product, no communication at all?

**Research Question no 4 (based on the (2)(4)):** Which of the created combination of products in multiproduct offer represents the higher response rate from the given product perspective in the particular segments?

Moreover, there is very useful, analytically desirable and noteworthy information on the created segments and the verification of the main offer dedicated to the particular segment. That is the reason for the last question presence.

**Research Question no 5 (analytical):** How do the segments work? Is the expected response rate similar to those obtained in the campaign? Is the dedicated offer (from the Champion groups) the most suitable? (Segmentation assumption verification).

Besides, the last but also the most obvious one – it could be done by the comparison of the response rates of particular bank products from the whole campaign perspective (1) and segment perspective (2). Therefore, the last research question is:

**Research Question no 6**: What are the response rates of all products offered in the whole campaign and in the particular segments?

In the next part of this section each Research Question will be developed by finding the numbers and formulated the answers. Detailed data about MPO test results are to be found in Appendix D.

### 5.2.1 Research Question no 1

In the context of the bank where the research was taken it is worth mentioning, that Credit Product no 1 has been (and also still keeps this 'position') the one with the biggest interest of product managers to be sold and also the most popular bank product among the customers. Because of this, the MPO approach is treated with a particular attention by this manager.

The Research Question no 1 (**RQ1**) investigates how high response rates are obtained in the whole campaign (MPO test) for each of four proposed bank products. Is there any difference between response rates achieved in the groups of customers with different types of communication: single-product offering, multiproduct offering with given product, any other communication without given product or no communication at all?

In practice, starting from the CP1 view, it is necessary to find the response rate for the CP1 in the whole campaign, particularly in the groups where single CP1 were offered. Then, find the response rate for the CP1 in the whole campaign, in the groups where CP1 were offered together with other products. As the next step, find the response rate for the CP1 in the whole campaign in the groups with other type of communication, where customer is contacted by bank but without CP1 offer. The last step is to find the response rate for the CP1 in the whole campaign, in the groups without any offer, without any contact from bank. All these steps are similarly applied to the CP2, CP3 and DP1.

The table below shows results for all four products in the order mentioned above.

*Table 5.2.1 Response rates as the answers for Research Question no 1.*

| Perspective (1) | Perspective (3) | Product | Response rate |
|---|---|---|---|
| the whole MPO campaign | single product offer | CP1 | 2,81% |
| the whole MPO campaign | multiproduct offer | CP1 | 2,71% |
| the whole MPO campaign | other offer (not given product) | CP1 | 1,64% |
| the whole MPO campaign | any offer | CP1 | 1,37% |
| the whole MPO campaign | single product offer | CP2 | 0,97% |
| the whole MPO campaign | multiproduct offer | CP2 | 0,93% |
| the whole MPO campaign | other offer (not given product) | CP2 | 0,41% |
| the whole MPO campaign | any offer | CP2 | 0,49% |
| the whole MPO campaign | single product offer | CP3 | 0,55% |
| the whole MPO campaign | multiproduct offer | CP3 | 0,44% |
| the whole MPO campaign | other offer (not given product) | CP3 | 0,44% |
| the whole MPO campaign | any offer | CP3 | 0,29% |
| the whole MPO campaign | single product offer | DP1 | 0,64% |
| the whole MPO campaign | multiproduct offer | DP1 | 1,30% |
| the whole MPO campaign | other offer (not given product) | DP1 | 1,88% |
| the whole MPO campaign | any offer | DP1 | 0,62% |

For **CP1** the biggest response rate is noted when the customer is offered single CP1 offer (in all possibilities) and the lowest response rate is in the group without any offer. The MPO gives the second result, but the difference between these two results (a single-offer and MPO) is not statistically significant.

For **CP2** situation is the same – the highest response rate is noted when customer is offered single CP2. But when CP2 is a part of MPO the result is almost as high as for single offering and the difference between these two results is not statistically significant. The lowest response rate is noted for CP2 when the customer is contacted by bank with other offer (not CP2). It means that the customers, who are not offered by bank, buy CP2 more often than those customers, who are not contacted by bank. Although it is worth adding that the difference between those two last results is not statistically significant neither.

The last customer finance product – **CP3** gains the highest result in the groups where CP3 was offered as a single item. However, opposite to CP2 difference between MPO offering with CP3 as a part and offering customer with other products (single offering) is statistically significant. The lowest response rate is noted in the groups with no offer. It means that the natural need to buy CP3 by customer is less possible than need to buy CP3 which was stimulated by contact from the bank.

With **DP1** it is slightly different because the highest response rate is noted when customer is offered with any letter and the lowest when customer is not offered at all or when customer is offered with single DP1. It can mean that DP1 is not a product which is desired as the first need of the customer. But if customer is contacted by bank and gets any offer with any product, he/she realizes that he/she needs this proposed product.

### 5.2.2 Research Question no 2

Research Question no 2 (**RQ2**) investigates which of the created combination of products in multiproduct offer represents the higher response rate from the given product perspective in the whole campaign.
In practice, starting from CP1 view as well to compare the results is necessary to count response rate for proposed product combination, meaning (CP1|CP2|CP3) (i), (CP1|CP2|DP1) (ii), (CP1|CP2) (iii), (DP1|CP3) (iv). In (i), (ii), (iii) group CP1 is offered by the bank, but in the last one (iv), CP1 is not a part of the offer. For the remaining products there are also the same calculations done, with this difference that CP2 is a part of (i), (ii), (iii), CP3 is a part of (iii) and (iv) combination and DP1 is also a part of (iii) and (iv) multiproduct offer. Obtained results are presented in the table below.

*Table 5.2.2 Response rates as the answers for Research Question no 2.*

| Perspective (1) | Perspective (4) | Product | Response rate |
|---|---|---|---|
| the whole MPO campaign | MPO: (CP1\|CP2\|CP3) (i) | CP1 | 3,35% |
| the whole MPO campaign | MPO: (CP1\|CP2\|DP1) (ii) | CP1 | 1,97% |
| the whole MPO campaign | MPO: (CP1\|CP2) (iii) | CP1 | 2,81% |
| the whole MPO campaign | MPO: (DP1\|CP3) (iv) | CP1 | 0,10% |
| the whole MPO campaign | MPO: (CP1\|CP2\|CP3) (i) | CP2 | 1,15% |
| the whole MPO campaign | MPO: (CP1\|CP2\|DP1) (ii) | CP2 | 0,54% |
| the whole MPO campaign | MPO: (CP1\|CP2) (iii) | CP2 | 1,12% |
| the whole MPO campaign | MPO: (DP1\|CP3) (iv) | CP2 | 0,10% |
| the whole MPO campaign | MPO: (CP1\|CP2\|CP3) (i) | CP3 | 0,57% |
| the whole MPO campaign | MPO: (CP1\|CP2\|DP1) (ii) | CP3 | 0,58% |
| the whole MPO campaign | MPO: (CP1\|CP2) (iii) | CP3 | 0,62% |
| the whole MPO campaign | MPO: (DP1\|CP3) (iv) | CP3 | 0,00% |
| the whole MPO campaign | MPO: (CP1\|CP2\|CP3) (i) | DP1 | 1,21% |
| the whole MPO campaign | MPO: (CP1\|CP2\|DP1) (ii) | DP1 | 1,64% |
| the whole MPO campaign | MPO: (CP1\|CP2) (iii) | DP1 | 0,64% |
| the whole MPO campaign | MPO: (DP1\|CP3) (iv) | DP1 | 0,13% |

As far as **CP1** is concerned, the highest response rate is noted for the typical multiproduct offer for consumer finance customers. It consists of all the credit products (CP1, CP2 and CP3). It means that getting the information about the wide bank credit portfolio can stimulate customer need. A double offer with CP1 and CP2 is on the second place with its result. It also means that combination of these two offer stimulates customer to buy more CP1. However, the difference between these two response rates (MPO and double) is statistically significant, what can be summarized that the more credit offer is put in one letter, the higher response rate is noted for CP1. And when credit products are combined with DP1 the response rate is the lowest.

As far as **CP2** is concerned in the context of RQ2, the results are similar to the results for CP1: the highest response rate is noted for MPO (CP1|CP2|CP3) combination. However, the results are simultaneoulsy different because the difference in response rates between (i) offering and (iii) offering is not statistically significant. It concludes that there is no difference if CP2 is combined with only CP1 or with CP1 and CP3 because obtained results are almost the same. When combining credit products with DP1 the situation is the same: the response rate is like more than 50% lower.

For **CP3** there is no statistically significant difference between multiproduct groups. However, the highest response rate is noted in the double group (iii). It seems strange as in this combination there is no CP3 offer available. It could mean that CP3 is the product which is purchased by the customers who have got the real need to buy this actual product and hence, with high CP3 awareness (bought, even if not communicated).

For **DP1** the highest response rate is noted for (ii) offering. It means that when customer is getting DP1 offer supported by credit products it can ease DP1 buying awareness. On the other hand, when DP1 is combined with CP3 (also product from the credit side), the response rate is the lowest. It can mean that the customer`s need of buying DP1 is directly correlated with the need of buying CP1 and CP2 and inversely correlated with the need of buying CP3.

### 5.2.3 Research Question no 3

Another Research Question (**RQ3**) investigates similar issues as in RQ1, but in the segments context instead of the whole bank perspective. In fact it concerns the question about how high response rates are obtained in the particular segments for each of four proposed bank products. Is there any difference between response rates obtained in the groups of customers with different types of communication: single-product offering, multiproduct offering with given product, any other communication without given product or no communication at all? All obtained results are cumulated in the table below.

*Table 5.2.3 Response rates as the answers for Research Question no 3.*

| Perspective (2) | Perspective (3) | Segment | Response rate CP1 | Response rate CP2 | Response rate CP3 | Response rate DP1 |
|---|---|---|---|---|---|---|
| segment perspective | single product offer | Segment no 1 | 3,46% | 0,83% | 0,35% | 0,62% |
| segment perspective | multiproduct offer | Segment no 1 | 2,92% | 0,96% | 0,28% | 0,42% |
| segment perspective | other offer (not given product) | Segment no 1 | 2,24% | 0,42% | 0,31% | 0,66% |
| segment perspective | no offer | Segment no 1 | 2,08% | 0,48% | 0,42% | 0,48% |
| segment perspective | single product offer | Segment no 2 | 3,85% | 2,32% | 1,63% | 1,76% |
| segment perspective | multiproduct offer | Segment no 2 | 3,07% | 1,14% | 0,73% | 3,31% |
| segment perspective | other offer (not given product) | Segment no 2 | 2,70% | 0,67% | 0,97% | 0,10% |
| segment perspective | no offer | Segment no 2 | 3,95% | 1,69% | 0,75% | 2,07% |
| segment perspective | single product offer | Segment no 3 | 0,43% | 0,32% | 0,00% | 0,16% |
| segment perspective | multiproduct offer | Segment no 3 | 0,49% | 0,22% | 0,00% | 0,05% |
| segment perspective | other offer (not given product) | Segment no 3 | 0,23% | 0,10% | 0,04% | 0,10% |
| segment perspective | no offer | Segment no 3 | 0,28% | 0,07% | 0,14% | 0,14% |
| segment perspective | single product offer | Segment no 4 | 5,80% | 1,17% | 0,79% | 1,39% |
| segment perspective | multiproduct offer | Segment no 4 | 5,00% | 1,61% | 1,00% | 1,08% |
| segment perspective | other offer (not given product) | Segment no 4 | 4,62% | 1,21% | 0,83% | 1,75% |
| segment perspective | no offer | Segment no 4 | 2,65% | 0,99% | 0,66% | 1,49% |

| | | | | | | |
|---|---|---|---|---|---|---|
| segment perspective | single product offer | Segment no 5 | 0,52% | 0,22% | 0,00% | 0,16% |
| segment perspective | multiproduct offer | Segment no 5 | 0,25% | 0,18% | 0,00% | 0,14% |
| segment perspective | other offer (not given product) | Segment no 5 | 0,23% | 0,08% | 0,06% | 0,13% |
| segment perspective | no offer | Segment no 5 | 0,18% | 0,05% | 0,00% | 0,03% |

For **CP1** and Segment no 1 the highest response rate is noted for single CP1 offer. In Segment no 2 the highest response rate is observed in the group with no offer. At the same time the difference between this response rate and the response rate in single product offer is not statistically significant. In Segment no 3 all the response rates achieved are quite low, but the highest result is realized with multiproduct offer. Segment no 4 is characterized by the highest value of response rates and the best one is in the single product offer group. In Segment no 5 response rates are the same low as in Segment no 3, but still the highest response rate is noted in the group with single product offer.

Response rate of **CP2** in Segment no 1 is the highest in the MPO group. It can be translated that CP2 is complementary product rather than the main. In Segment no 2 results are the best comparing to the rest of segments and the highest score is represented by the group with single product offer. Segment no 3 is characterized by relatively low response rates, but the highest results is noted for the group with single product offer. In Segment no 4, the same as in Segment no 2, the highest response rate is noted for MPO. Therefore, it allows to assume that Segment no 4 is similar to Segment no 2 from the CP2 point of view. Segment no 5 represents quite low results, as Segment no 3, but the highest response rate is put for the single product offer.

For **CP3** and its behavior in Segment no 1, the highest response rate is observed for offers which are not including the CP3 proposition. It could mean that the customers are not coming to the bank to buy CP3. This product is chosen with the rule based on the similarity to the credit products. CP3 has almost no results in Segment no 3, although the highest result is also for the group where customers are not offered. In Segment no 4 the highest response rate is noted for MPO. In Segment no 5 it could be assumed that there are no results for the customers who get CP3 offer, so if the customers are not likely to buy CP3, there is no chance to buy it in any other product configuration.

As far as **DP1** is concerned, in Segment no 1 there is similar situation as for CP3, meaning that the highest response rate is presented for group of customers who are offered with any other products, but not including DP1. In Segment no 2 the best results are obtained for MPO. MPO is the offer dedicated to this segment. In Segment no 3, which presents quite low results, the highest one is noted for the group with single product offer. In Segment no 4 there is again the same situation as for CP3: when customers are offered any other product, but no DP1, the response rate is the highest. Segment no 5 cumulates customers with very low propensity-to-buy for all products.

### 5.2.4 Research Question no 4

Research Question no 4 (RQ4) in general considers the same issues as RQ2, but from the created segments perspective. It investigates which of the created combination of products in multiproduct offer represents the higher response rate from the given product perspective in the particular segments.

*Table 5.2.4 Response rates as the answers for Research Question no 4.*

| Perspective (2) | Perspective (3) | Segment | Respon se rate CP1 | Respon se rate CP2 | Respon se rate CP3 | Respon se rate DP1 |
|---|---|---|---|---|---|---|
| segment perspective | MPO: (CP1\|CP2\|CP3) (i) | Segment no 1 | 3,25% | 0,90% | 0,28% | 0,55% |
| segment perspective | MPO: (CP1\|CP2\|DP1) (ii) | Segment no 1 | 2,01% | 0,55% | 0,21% | 0,42% |
| segment perspective | MPO: (CP1\|CP2) (iii) | Segment no 1 | 3,12% | 1,12% | 0,43% | 0,74% |

| | | | | | | |
|---|---|---|---|---|---|---|
| segment perspective | MPO: (DP1\|CP3) (iv) | Segment no 1 | 0,00% | 0,00% | 0,00% | 0,00% |
| segment perspective | MPO: (CP1\|CP2\|CP3) (i) | Segment no 2 | 3,11% | 1,28% | 0,73% | 2,92% |
| segment perspective | MPO: (CP1\|CP2\|DP1) (ii) | Segment no 2 | 2,78% | 0,66% | 1,06% | 3,31% |
| segment perspective | MPO: (CP1\|CP2) (iii) | Segment no 2 | 3,86% | 2,39% | 1,86% | 0,76% |
| segment perspective | MPO: (DP1\|CP3) (iv) | Segment no 2 | 0,00% | 0,00% | 0,00% | 0,00% |
| segment perspective | MPO: (CP1\|CP2\|CP3) (i) | Segment no 3 | 0,43% | 0,21% | 0,00% | 0,11% |
| segment perspective | MPO: (CP1\|CP2\|DP1) (ii) | Segment no 3 | 0,51% | 0,31% | 0,10% | 0,00% |
| segment perspective | MPO: (CP1\|CP2) (iii) | Segment no 3 | 0,54% | 0,15% | 0,07% | 0,07% |
| segment perspective | MPO: (DP1\|CP3) (iv) | Segment no 3 | 0,00% | 0,00% | 0,00% | 0,16% |
| segment perspective | MPO: (CP1\|CP2\|CP3) (i) | Segment no 4 | 5,58% | 1,88% | 1,00% | 1,57% |
| segment perspective | MPO: (CP1\|CP2\|DP1) (ii) | Segment no 4 | 2,76% | 0,87% | 0,59% | 1,08% |
| segment perspective | MPO: (CP1\|CP2) (iii) | Segment no 4 | 5,57% | 1,60% | 1,02% | 1,38% |
| segment perspective | MPO: (DP1\|CP3) (iv) | Segment no 4 | 0,00% | 0,00% | 0,00% | 0,00% |
| segment perspective | MPO: (CP1\|CP2\|CP3) (i) | Segment no 5 | 0,13% | 0,09% | 0,00% | 0,18% |
| segment perspective | MPO: (CP1\|CP2\|DP1) (ii) | Segment no 5 | 0,22% | 0,09% | 0,04% | 0,18% |
| segment perspective | MPO: (CP1\|CP2) (iii) | Segment no 5 | 0,40% | 0,35% | 0,08% | 0,07% |
| segment perspective | MPO: (DP1\|CP3) (iv) | Segment no 5 | 0,19% | 0,19% | 0,00% | 0,10% |

For **CP1** the highest response rate in Segment no 1 is obtained for the offer of consumer finance products, meaning configuration of all three credit products combining in one bank offer. In Segment no 2 the highest response rate is noted in the group offered two credit products – CP1 and CP2. In Segment no 3 the highest response rate comes from the double offer. However, group with mixed offer combining two credit products and one saving product (CP1, CP2 and DP1) is with no statistically significant difference. In Segment no 4 there is no practical difference between full credit offer (i) and double credit offer (iii). It could mean that from CP1 point of view there is no difference if offer includes only CP1 and CP2 or CP3 as well. InSegment no 5 the highest response rate (but still very low) is noted for double credit offer (iii).

For **CP2** the highest response rate in Segment no 1, Segment no 2 and Segment no 5 is noted for double credit offer (iii). In Segment no 3 offer which includes CP1, CP2 and DP1 (ii) represents the highest results. In Segment no 4 the highest Response rate is noted for MPO consists of credit products (i).
For **CP3** and Segment no 1, Segment no 2 and Segment no 4 the highest Response rate is noted for double offer (iii). However, in Segment no 4 the almost the same high response rate is noted for MPO (i). Segment no 3 and Segment no 5 show relatively low response rates for CP3, but in Segment no 3 the higher response rate is get for MPO included DP1 (ii) and in Segment no 5 the higher response rate is noted in group with double offer (iii).
As far as **DP1** is concerned, in Segment no 1 the customers with the highest response are offered with double credit offer (iii). In Segment no 2, the customers from the group of MPO included DP1 (ii) let obtain the highest response rate from the campaign. For Segment no 3, that represents almost the lowest response rates, the higher one is achieved in the group with other offers than DP1. Segment no 4 shows the highest response rate in the group with MPO with credit products (i). In Segment no 5 the highest response rate is noted in two MPOs – all credit products offer (i) and double credit with DP1 offer (ii).

### 5.2.5 Research Question no 5

Typical analytical Research Question no 5 (**RQ5**) investigates how the segments work. Is the expected response rate similar to those really obtained? Is the dedicated offer (from the Champion groups) the most suitable? In practice this investigation should show, if customers who are concentrated in a given cluster are given the most appropriate offer according to their propensity-to-buy estimations. Results are shown in the table below.

*Table 5.2.5.1 Response rates as the answers for Research Question no 5, for products bought after communication.*

| RESULTS OF PRODUCT BOUGHT AFTER COMMUNICATION | | | | | |
|---|---|---|---|---|---|
| Perspective | Segment | Response rate CP1 | Response rate CP2 | Response rate CP3 | Response rate DP1 |
| (2) | S1_CL_01 / No offer | 0,00% | 0,00% | 0,00% | 0,00% |
| (2) | S1_CL_02 / CP2 | 0,00% | 0,86% | 0,00% | 0,00% |
| (2) | S1_CL_03 / DP1 | 0,00% | 0,00% | 0,00% | 0,65% |
| (2) | S1_CL_04 / CP3 | 0,00% | 0,00% | 0,35% | 0,00% |
| (2) | S1_CL_05 / CP1\|CP2\|CP3 | 3,23% | 0,87% | 0,27% | 0,00% |
| (2) | S1_CL_06 / CP1\|CP2\|DP1 | 1,97% | 0,52% | 0,00% | 0,45% |
| (2) | S1_CL_07 / CP1 | 3,46% | 0,00% | 0,00% | 0,00% |
| (2) | S1_CM_00 / CP1\|CP2 | 3,12% | 1,13% | 0,00% | 0,00% |
| (2) | S2_CL_01 / CP1 | 3,85% | 0,00% | 0,00% | 0,00% |
| (2) | S2_CL_02 / CP2 | 0,00% | 2,32% | 0,00% | 0,00% |
| (2) | S2_CL_03 / DP1 | 0,00% | 0,00% | 0,00% | 1,76% |
| (2) | S2_CL_04 / CP3 | 0,00% | 0,00% | 1,63% | 0,00% |
| (2) | S2_CL_05 / CP1\|CP2\|CP3 | 3,11% | 1,28% | 0,73% | 0,00% |
| (2) | S2_CL_06 / No offer | 0,00% | 0,00% | 0,00% | 0,00% |
| (2) | S2_CL_07 / CP1\|CP2 | 3,86% | 2,39% | 0,00% | 0,00% |
| (2) | S2_CM_00 / CP1\|CP2\|DP1 | 2,78% | 0,66% | 0,00% | 3,31% |
| (2) | S3_CL_01 / CP1 | 0,43% | 0,00% | 0,00% | 0,00% |
| (2) | S3_CL_02 / CP2 | 0,00% | 0,32% | 0,00% | 0,00% |
| (2) | S3_CL_03 / No offer | 0,00% | 0,00% | 0,00% | 0,00% |
| (2) | S3_CL_04 / CP3 | 0,00% | 0,00% | 0,00% | 0,00% |
| (2) | S3_CL_05 / CP1\|CP2\|CP3 | 0,43% | 0,21% | 0,00% | 0,00% |
| (2) | S3_CL_06 / CP1\|CP2\|DP1 | 0,51% | 0,31% | 0,00% | 0,00% |
| (2) | S3_CL_07 / DP1\|CP3 | 0,00% | 0,00% | 0,00% | 0,16% |
| (2) | S3_CL_08 / CP1\|CP2 | 0,54% | 0,15% | 0,00% | 0,00% |
| (2) | S3_CM_00 / DP1 | 0,00% | 0,00% | 0,00% | 0,16% |
| (2) | S4_CL_01 / CP1 | 5,80% | 0,00% | 0,00% | 0,00% |
| (2) | S4_CL_02 / CP2 | 0,00% | 1,17% | 0,00% | 0,00% |
| (2) | S4_CL_03 / DP1 | 0,00% | 0,00% | 0,00% | 1,39% |
| (2) | S4_CL_04 / CP3 | 0,00% | 0,00% | 0,79% | 0,00% |
| (2) | S4_CL_05 / No offer | 0,00% | 0,00% | 0,00% | 0,00% |
| (2) | S4_CL_06 / CP1\|CP2\|DP1 | 2,76% | 0,87% | 0,00% | 1,08% |
| (2) | S4_CL_07 / CP1\|CP2 | 5,57% | 1,60% | 0,00% | 0,00% |
| (2) | S4_CM_00 / CP1\|CP2\|CP3 | 5,58% | 1,88% | 1,00% | 0,00% |
| (2) | S5_CL_01 / CP1 | 0,52% | 0,00% | 0,00% | 0,00% |
| (2) | S5_CL_02 / CP2 | 0,00% | 0,22% | 0,00% | 0,00% |
| (2) | S5_CL_03 / DP1 | 0,00% | 0,00% | 0,00% | 0,13% |
| (2) | S5_CL_04 / CP3 | 0,00% | 0,00% | 0,00% | 0,00% |
| (2) | S5_CL_05 / CP1\|CP2\|CP3 | 0,13% | 0,09% | 0,00% | 0,00% |
| (2) | S5_CL_06 / CP1\|CP2\|DP1 | 0,22% | 0,09% | 0,00% | 0,18% |
| (2) | S5_CL_07 / DP1\|CP3 | 0,00% | 0,00% | 0,00% | 0,10% |
| (2) | S5_CL_08 / CP1\|CP2 | 0,40% | 0,35% | 0,00% | 0,00% |
| (2) | S5_CM_00 / No offer | 0,00% | 0,00% | 0,00% | 0,00% |

This comparison could be conducted from two sides. First, it can start from the comparing only these results, where the customer bought the concrete products which appear in the offer (like in the Table 5.2.5.1). In Segment no 1, which focuses on the customers with the highest propensity-to-buy of CP1 and CP2, Champion group for the single CP1 offer does not show the highest response rate for this product. However, the response rate for CP2 is the highest. In Segment no 2, which cumulates customers with high propensity-to-buy of CP1, CP2 and DP1, response rates in champion group for CP1 and CP2 are not the highest, but for DP1 are the best. Segment no 3 was assumed to focus on customers likely to buy DP1. Meanwhile, because of the bank strategy and very long consumer finance history of customers behavior the real and actual DP1 purchases are relatively low. But if DP1 is offered as a single item, or in the cooperation with CP3, response rate is the same. Segment no 4 was expected to focus on customers who are very likely to buy credit products (CP1, CP2, CP3). Although response rate in MPO offer for CP1 is not the highest (but it is quite high), for CP2 and CP3 it is the best out of all noted responses. Segment no 5 was described as not likely to buy any product and in fact, when customers from this segment are not offered, the response rate is equal 0%.

*Table 5.2.5.2 Response rates as the answers for Research Question no 5, for all products.*

| Perspective | Segment | Response rate CP1 | Response rate CP2 | Response rate CP3 | Response rate DP1 |
|---|---|---|---|---|---|
| (2) | S1_CL_01 / No offer | 2,08% | 0,48% | 0,42% | 0,48% |
| (2) | S1_CL_02 / CP2 | 2,98% | 0,83% | 0,28% | 0,55% |
| (2) | S1_CL_03 / DP1 | 1,73% | 0,55% | 0,07% | 0,62% |
| (2) | S1_CL_04 / CP3 | 2,01% | 0,21% | 0,35% | 0,48% |
| (2) | S1_CL_05 / CP1\|CP2\|CP3 | 3,25% | 0,90% | 0,28% | 0,55% |
| (2) | S1_CL_06 / CP1\|CP2\|DP1 | 2,01% | 0,55% | 0,21% | 0,42% |
| (2) | S1_CL_07 / CP1 | 3,46% | 0,48% | 0,35% | 0,83% |
| (2) | S1_CM_00 / CP1\|CP2 | 3,12% | 1,12% | 0,43% | 0,74% |
| (2) | S2_CL_01 / CP1 | 3,85% | 0,22% | 0,66% | 2,64% |
| (2) | S2_CL_02 / CP2 | 4,01% | 2,32% | 0,63% | 5,28% |
| (2) | S2_CL_03 / DP1 | 1,44% | 0,16% | 0,48% | 1,76% |
| (2) | S2_CL_04 / CP3 | 2,66% | 1,63% | 1,63% | 2,45% |
| (2) | S2_CL_05 / CP1\|CP2\|CP3 | 3,11% | 1,28% | 0,73% | 2,92% |
| (2) | S2_CL_06 / No offer | 3,95% | 1,69% | 0,75% | 2,07% |
| (2) | S2_CL_07 / CP1\|CP2 | 3,86% | 2,39% | 1,86% | 0,76% |
| (2) | S2_CM_00 / CP1\|CP2\|DP1 | 2,78% | 0,66% | 1,06% | 3,31% |
| (2) | S3_CL_01 / CP1 | 0,43% | 0,00% | 0,00% | 0,22% |
| (2) | S3_CL_02 / CP2 | 0,32% | 0,32% | 0,11% | 0,11% |
| (2) | S3_CL_03 / No offer | 0,28% | 0,07% | 0,14% | 0,14% |
| (2) | S3_CL_04 / CP3 | 0,43% | 0,00% | 0,00% | 0,00% |

| | | | | | |
|---|---|---|---|---|---|
| (2) | S3_CL_05 / CP1\|CP2\|CP3 | 0,43% | 0,21% | 0,00% | 0,11% |
| (2) | S3_CL_06 / CP1\|CP2\|DP1 | 0,51% | 0,31% | 0,10% | 0,00% |
| (2) | S3_CL_07 / DP1\|CP3 | 0,00% | 0,00% | 0,00% | 0,16% |
| (2) | S3_CL_08 / CP1\|CP2 | 0,54% | 0,15% | 0,07% | 0,07% |
| (2) | S3_CM_00 / DP1 | 0,21% | 0,21% | 0,00% | 0,16% |
| (2) | S4_CL_01 / CP1 | 5,80% | 1,19% | 1,02% | 2,22% |
| (2) | S4_CL_02 / CP2 | 5,24% | 1,17% | 0,72% | 1,90% |
| (2) | S4_CL_03 / DP1 | 4,26% | 1,48% | 0,78% | 1,39% |
| (2) | S4_CL_04 / CP3 | 4,37% | 0,96% | 0,79% | 2,01% |
| (2) | S4_CL_05 / No offer | 2,65% | 0,99% | 0,66% | 1,49% |
| (2) | S4_CL_06 / CP1\|CP2\|DP1 | 2,76% | 0,87% | 0,59% | 1,08% |
| (2) | S4_CL_07 / CP1\|CP2 | 5,57% | 1,60% | 1,02% | 1,38% |
| (2) | S4_CM_00 / CP1\|CP2\|CP3 | 5,58% | 1,88% | 1,00% | 1,57% |
| (2) | S5_CL_01 / CP1 | 0,52% | 0,05% | 0,09% | 0,24% |
| (2) | S5_CL_02 / CP2 | 0,22% | 0,22% | 0,09% | 0,09% |
| (2) | S5_CL_03 / DP1 | 0,24% | 0,03% | 0,00% | 0,16% |
| (2) | S5_CL_04 / CP3 | 0,27% | 0,05% | 0,00% | 0,09% |
| (2) | S5_CL_05 / CP1\|CP2\|CP3 | 0,13% | 0,09% | 0,00% | 0,18% |
| (2) | S5_CL_06 / CP1\|CP2\|DP1 | 0,22% | 0,09% | 0,04% | 0,18% |
| (2) | S5_CL_07 / DP1\|CP3 | 0,19% | 0,19% | 0,00% | 0,10% |
| (2) | S5_CL_08 / CP1\|CP2 | 0,40% | 0,35% | 0,08% | 0,07% |
| (2) | S5_CM_00 / No offer | 0,18% | 0,05% | 0,00% | 0,03% |

Secondly, as shown in Table 5.2.5.2, the customer could buy bank products regardless of the offer he/she got. Based on this possibility, in Segment no 1 for CP2 results achieved in champion group is still the highest one. Moreover, in this group there is also the highest response rate achieved by CP3, even if this product is not the leading one in this segment. Response rate of CP1 is the highest for single product offer group and then for MPO credit group. To get the most optimal results, which are the most suitable to the Segment, offer should be construct according to the champion group. In Segment no 2 results obtained in champion groups are not the highest for any product. Moreover, there is no clear rule how change the approach if needed because the response rate for CP1 is the highest but while offering the single CP2. The same is noted for DP1. For CP2 and CP3 the highest response rate is noted for group with double offer. Simultaneously, quite high response rates are read for group with no offer. It allows to assume that this Segment consists of natural buyers, meaning customers of high product and need awareness, who do not need any letter from the bank to buy bank products.

Segment no 3 presents relatively low results, also for leading DP1. The best activity for this Segment is not sending them offer and wait for the change of the strategy.

Segment no 4 was expressed as being the most consumer finance because it focuses on the customers who are the most likely to buy all the credit products. It presents relatively the highest response rates for all products comparing to remaining Segments. Champion group shows the highest response rate for CP2 and CP3. Also for CP1 response rate in this group is quite high, even not the highest.

Segment no 5 was rated as full of customers who are not likely to buy bank products. And even if some customers bought CP1, a great majority of the Segment did not buy anything.

For more details and also the financial side of the created segments, please read Appendix D.

### 5.2.6 Research Question no 6

The last Research Question (**RQ6**) simply investigates what the response rates of all products offered in the whole campaign and in the particular segments are. In order to achieve these results it is necessary to simply count response rates for all particular segments and then for the whole MPO test. The results are to be found below.

*Table 5.2.6 Response rates as the answers for Research Question no 6.*

| Perspective | | Product | Response rate |
|---|---|---|---|
| (2) | SEGMENT no 1 | CP1 | 2,68% |
| (2) | SEGMENT no 2 | CP1 | 3,12% |
| (2) | SEGMENT no 3 | CP1 | 0,33% |
| (2) | SEGMENT no 4 | CP1 | 4,73% |
| (2) | SEGMENT no 5 | CP1 | 0,25% |
| (1) | the whole MPO campaign | CP1 | 1,70% |
| (2) | SEGMENT no 1 | CP2 | 0,73% |
| (2) | SEGMENT no 2 | CP2 | 1,17% |
| (2) | SEGMENT no 3 | CP2 | 0,15% |
| (2) | SEGMENT no 4 | CP2 | 1,38% |
| (2) | SEGMENT no 5 | CP2 | 0,11% |

| | | | |
|---|---|---|---|
| (1) | the whole MPO campaign | CP2 | 0,69% |
| (2) | SEGMENT no 1 | CP3 | 0,32% |
| (2) | SEGMENT no 2 | CP3 | 0,99% |
| (2) | SEGMENT no 3 | CP3 | 0,04% |
| (2) | SEGMENT no 4 | CP3 | 0,86% |
| (2) | SEGMENT no 5 | CP3 | 0,03% |
| (1) | the whole MPO campaign | CP3 | 0,43% |
| (2) | SEGMENT no 1 | DP1 | 0,61% |
| (2) | SEGMENT no 2 | DP1 | 2,78% |
| (2) | SEGMENT no 3 | DP1 | 0,12% |
| (2) | SEGMENT no 4 | DP1 | 1,62% |
| (2) | SEGMENT no 5 | DP1 | 0,11% |
| (1) | the whole MPO campaign | DP1 | 1,01% |

CP1, as the informal main bank product in the research context got the highest response rate in the whole MPO test campaign. From the segments perspective it is Segment no 4 which achieved the highest respose rate for CP1. Segment no 4 was defined as multiproduct one according to the propensity-to-buy scores. It consists of the customers who are very likely to buy credit products. The lowest response rate is noted for Segment no 5 which was defined as the No offer group and includes customers with very low probability to buy any bank product. Also low results are noted in Segment no 3 which is offered with DP1.

CP2, which is less popular product than CP1, represents the third in the order result of the response rate in the whole MPO test campaign. From the segments perspective conclusions are similar to those for CP1. The highest response rate is noted for Segment no 4, then for Segment no 2. The lowest results are given by Segment no 5 and Segment no 3.

CP3, with the lowest response rate in the whole MPO test campaign achieves the highest response rate in Segment no 2, then in Segment no 4. Segment no 5 and Segment no 3 show almost the same low response rate.

DP1, the product with the second result of response rate in the whole MPO test campaign also concludes with the highest response rate in Segment no 2, then in Segment no 4. The lowest response rate is noted in Segment no 5, but response rate in Segment no 3 is the same low. It could be perceived as quite strange result since Segment no 3 was defined as likely to buy DP1 mostly. The reason of such situation could be the fact that this segment includes the customers with only consumer finance history. Therefore, even if propensity-to-buy scores for these customers are high and they are qualified into Decile 1 these customers are most of all likely to buy credit products. Another reason of such a low resuts is a bank strategy which after financial crisis is far from collecting deposits.

### 5.2.7 Which approach gives the better results? Theoretically estimation.

In a customer life time cycle a once acquired customer has to be retained using many active management processes. One of these processes described in Chapter 2, is cross-sell action. The most effective from an optimization point of view is offering the right product to right customer since such an activity is going to save the money on targeting only the proper customers who are defined as the most likely to buy proper bank products. Each customer is described with the propensity-to-buy scores of four key bank products: CP1, CP2, CP3, DP1. Next to the MPO test **single-product cross-sell campaign** is also carried out. Its results are as follow:

*Table 5.2.7.1 Direct CRM results of the single-product campaigns.*

| | | PRODUCT COMMUNICATED AND BOUGHT | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | CP1 | | CP2 | | CP3 | | DP1 | |
| OFFER TYPE | SIZE | RR NO | RR % | RR NO | RR % | RR NO | RR % | RR NO | RR % |
| CP1 | 1 004 740 | **26 236** | **2,61%** | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% |
| CP2 | 298 180 | 0 | 0,00% | **1 573** | **0,53%** | 0 | 0,00% | 0 | 0,00% |
| CP3 | 445 362 | 0 | 0,00% | 0 | 0,00% | **1 540** | **0,35%** | 0 | 0,00% |
| DP1 | 126 888 | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% | **16 231** | **12,79%** |

*Table 5.2.7.2 All the results of the single-product campaigns (regardless of the received offer).*

| | | PRODUCT BOUGHT EVEN IF NOT COMMUNICATED | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | CP1 | | CP2 | | CP3 | | DP1 | |
| CAMPAIGN NAME | SIZE | RR NO | RR % | RR NO | RR % | RR NO | RR % | RR NO | RR % |
| CP1 | 1 004 740 | **26 236** | **2,61%** | 4 729 | 0,47% | 2 948 | 0,29% | 5 337 | 0,53% |
| CP2 | 298 180 | 7 531 | 2,53% | 1 573 | 0,53% | 542 | 0,18% | 1 497 | 0,50% |
| CP3 | 445 362 | 8 017 | 1,80% | 1 870 | 0,42% | 1 540 | 0,35% | 1 972 | 0,44% |
| DP1 | 126 888 | 2 526 | 1,99% | **1 912** | **1,51%** | **2 071** | **1,63%** | **16 231** | **12,79%** |

CP1 offer gives the biggest number of sold products in number. The second campaign is the one with DP1 offer. Next, there is a campaign with CP2 offer. The smallest sales are noted in campaign with CP3 offer. CP1 offer, because of its size, gives the biggest number of sales of CP1. In this campaign there is also the biggest number of the other Credit Products sold, even if they were not communicated. However, if all the achieved response rates are taken into account, then this Campaign (with CP1 offer) takes the third place in the classification. CP2 gives the best results in Campaign with DP1 offer. The last product CP3 notes the highest response rate also in Campaign with DP1 and this Campaign shows the highest response rate of DP1 sale as well.

Such a presentation of the results does not allow to compare the existing single-product offering approach with proposed MPO approach. Because of this reason **the multiproduct clusters were assigned to each of the bank customer**. Afterwards, single-product campaigns were reported with using appropriate multiproduct segment. The modified results, after having the clusters assigned, are presented in the following table.

*Table 5.2.7.3 Direct CRM results of the single-product campaigns with division on 5 assigned multiproduct segments.*

| | | | | PRODUCT COMMUNICATED AND BOUGHT | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | CP1 | | CP2 | | CP3 | | DP1 | |
| NAME | OFFER TYPE | SEG | SIZE | RR NO | RR % | RR NO | RR % | RR NO | RR % | RR NO | RR % |
| CMPG_CP1 | CP1 | SEG_1 | 403 130 | 11 258 | 2,79% | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% |
| CMPG_CP1 | CP1 | SEG_2 | 75 450 | 2 051 | 2,72% | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% |
| CMPG_CP1 | CP1 | SEG_3 | 76 967 | 310 | 0,40% | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% |
| CMPG_CP1 | CP1 | SEG_4 | 325 467 | 12 227 | 3,76% | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% |
| CMPG_CP1 | CP1 | SEG_5 | 123 726 | 390 | 0,32% | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% |
| CMPG_CP2 | CP2 | SEG_1 | 146 768 | 0 | 0,00% | 723 | 0,49% | 0 | 0,00% | 0 | 0,00% |
| CMPG_CP2 | CP2 | SEG_2 | 31 194 | 0 | 0,00% | 250 | 0,80% | 0 | 0,00% | 0 | 0,00% |
| CMPG_CP2 | CP2 | SEG_3 | 24 194 | 0 | 0,00% | 19 | 0,08% | 0 | 0,00% | 0 | 0,00% |
| CMPG_CP2 | CP2 | SEG_4 | 56 397 | 0 | 0,00% | 540 | 0,96% | 0 | 0,00% | 0 | 0,00% |
| CMPG_CP2 | CP2 | SEG_5 | 39 627 | 0 | 0,00% | 41 | 0,10% | 0 | 0,00% | 0 | 0,00% |
| CMPG_CP3 | CP3 | SEG_1 | 141 673 | 0 | 0,00% | 0 | 0,00% | 452 | 0,32% | 0 | 0,00% |
| CMPG_CP3 | CP3 | SEG_2 | 22 941 | 0 | 0,00% | 0 | 0,00% | 326 | 1,42% | 0 | 0,00% |
| CMPG_CP3 | CP3 | SEG_3 | 42 470 | 0 | 0,00% | 0 | 0,00% | 13 | 0,03% | 0 | 0,00% |
| CMPG_CP3 | CP3 | SEG_4 | 171 385 | 0 | 0,00% | 0 | 0,00% | 734 | 0,43% | 0 | 0,00% |
| CMPG_CP3 | CP3 | SEG_5 | 66 892 | 0 | 0,00% | 0 | 0,00% | 16 | 0,02% | 0 | 0,00% |
| CMPG_DP1 | DP1 | SEG_1 | 7 550 | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% | 640 | 8,48% |
| CMPG_DP1 | DP1 | SEG_2 | 56 286 | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% | 8 732 | 15,51% |
| CMPG_DP1 | DP1 | SEG_3 | 323 | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% | 38 | 11,80% |
| CMPG_DP1 | DP1 | SEG_4 | 48 099 | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% | 4 935 | 10,26% |
| CMPG_DP1 | DP1 | SEG_5 | 14 629 | 0 | 0,00% | 0 | 0,00% | 0 | 0,00% | 1 886 | 12,89% |

Table 5.2.7.3 shows results for customers who have got the bank offer of the single-product and bought exactly this product, which was mentioned in received letter. Based on the table it can be said that for Campaign with CP1 as communicated offer, the best results are to be distinguished for customers with Segment no 4 assigned. In Campaign with CP2 offer, the higher response rate is also noted for customers who are assigned with Segment no 4. In the third Campaign with CP3 as a communicated offer, the best outcome can be available for customers with Segment no 2 assigned, but it is worth remembering that it was the smallest segment which can be pointed within this Campaign. And in the last Campaign with

DP1 as the communicated offer, the highest response rate is observed for customers with Segment no 2 assigned.

The bank customers, who got the offer from the bank, they also have bought other products from the bank portfolio. The results of this research are put in the table below.

*Table 5.2.7.4 All the results of the single-product campaigns with division on 5 assigned multiproduct segments (regardless of the received offer).*

| | | | | PRODUCT BOUGHT EVEN IF NOT COMMUNICATED | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | CP1 | | CP2 | | CP3 | | DP1 | |
| NAME | OFFER TYPE | SEG | SIZE | RR NO | RR % | RR NO | RR % | RR NO | RR % | RR NO | RR % |
| CMPG_CP1 | CP1 | SEG_1 | 403 130 | 11 258 | 2,79% | 1 913 | 0,47% | 938 | 0,23% | 1 723 | 0,43% |
| CMPG_CP1 | CP1 | SEG_2 | 75 450 | 2 051 | 2,72% | 639 | 0,85% | 654 | 0,87% | 1 733 | 2,30% |
| CMPG_CP1 | CP1 | SEG_3 | 76 967 | 310 | 0,40% | 58 | 0,08% | 18 | 0,02% | 42 | 0,06% |
| CMPG_CP1 | CP1 | SEG_4 | 325 467 | 12 227 | 3,76% | 1 956 | 0,60% | 1 301 | 0,40% | 1 791 | 0,55% |
| CMPG_CP1 | CP1 | SEG_5 | 123 726 | 390 | 0,32% | 163 | 0,13% | 37 | 0,03% | 48 | 0,04% |
| CMPG_CP2 | CP2 | SEG_1 | 146 768 | 3 794 | 2,59% | 723 | 0,49% | 205 | 0,14% | 486 | 0,33% |
| CMPG_CP2 | CP2 | SEG_2 | 31 194 | 819 | 2,62% | 250 | 0,80% | 144 | 0,46% | 559 | 1,79% |
| CMPG_CP2 | CP2 | SEG_3 | 24 194 | 101 | 0,42% | 19 | 0,08% | 4 | 0,02% | 13 | 0,05% |
| CMPG_CP2 | CP2 | SEG_4 | 56 397 | 2 689 | 4,77% | 540 | 0,96% | 177 | 0,31% | 426 | 0,75% |
| CMPG_CP2 | CP2 | SEG_5 | 39 627 | 129 | 0,33% | 41 | 0,10% | 12 | 0,03% | 14 | 0,03% |
| CMPG_CP3 | CP3 | SEG_1 | 141 673 | 2 696 | 1,90% | 699 | 0,49% | 452 | 0,32% | 583 | 0,41% |
| CMPG_CP3 | CP3 | SEG_2 | 22 941 | 591 | 2,58% | 207 | 0,90% | 326 | 1,42% | 658 | 2,87% |
| CMPG_CP3 | CP3 | SEG_3 | 42 470 | 147 | 0,35% | 29 | 0,07% | 13 | 0,03% | 22 | 0,05% |
| CMPG_CP3 | CP3 | SEG_4 | 171 385 | 4 409 | 2,57% | 833 | 0,49% | 734 | 0,43% | 682 | 0,40% |
| CMPG_CP3 | CP3 | SEG_5 | 66 892 | 173 | 0,26% | 102 | 0,15% | 16 | 0,02% | 26 | 0,04% |
| CMPG_DP1 | DP1 | SEG_1 | **7 550** | 403 | 5,34% | 142 | 1,88% | 158 | 2,09% | 640 | 8,48% |
| CMPG_DP1 | DP1 | SEG_2 | 56 286 | 602 | 1,07% | 916 | 1,63% | 893 | 1,59% | 8 732 | 15,51% |
| CMPG_DP1 | DP1 | SEG_3 | 323 | 15 | 4,80% | 11 | 3,54% | 4 | 1,22% | 38 | 11,80% |
| CMPG_DP1 | DP1 | SEG_4 | 48 099 | 1 413 | 2,94% | 838 | 1,74% | 1 012 | 2,10% | 4 935 | 10,26% |
| CMPG_DP1 | DP1 | SEG_5 | 14 629 | 93 | 0,64% | 4 | 0,03% | 5 | 0,04% | 1 886 | 12,89% |

**CP1** was sold with the highest response rate in the Campaign with DP1 offer and for customers with Segment no 1 assigned (CP1.1). The next, **CP2** is being observed to have the highest response rate in Campaign with DP1 offer and Segment no 3 assigned (CP2.1). To keep the logical way of concluding the third one value of response rate is represented by Campaign with CP3 offer and Segment no 2 assigned (CP2.3). These results give the

information that CP2 gives the best results for customers from Campaign with CP2 offer and with Segment no 2 assigned. For **CP3** the highest response rate is observed in Campaign with DP1 offer and with Segment no 4 assigned (CP3.1). The second result is noted for customers from Campaign with CP3 offer and Segment no 2 assigned (CP3.2). The last **DP1** gives the highest response rate in Campaign with DP1 and Segment no 2 assigned (DP1.1). The second result is noted for the Campaign with CP3 offer and Segment no 2 assigned (DP1.2).

The effectiveness of CRM operations could be presented not only in the form of response rates results but also from the financial point of view. In was assumed that financial perspective depends only on the two simply elements and also simply dependence between them. One component makes the profit side, where profit is expressed as the volume of products which were bought multiplied by relevant margin indicator. The second element is responsible of the cost side which is defined by the expenditures on the letters sent as a form of communication with customers.

The following Table (5.2.7.5) provides information about the volumes which were bought with a concrete product by given customer in particular assigned multiproduct segments and campaigns. Sales are connected with all purchased products regardless of the received offer. Value of the average ticket in a given group that is defined by concrete campaign and segment are also counted. All values of volumes and average tickets should be multiple by one thousand to become a real values, but they are expressed in currency units adopted specifically for this research. Every campaign is summarized by summing up the bought products and volumes, by counting the response rates and average tickets for every campaign but with keeping the division into the sort of the product.

*Table 5.2.7.5 Summary of Response rates and Volumes achieved in single-product campaigns (regardless of the received offer) in division on assigned segments from MPO approach and on bank products, regardless of the received offer.*

| | | CP1 | | | | CP2 | | | | CP3 | | | | DP1 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NAME | SEG | RR NO | RR % | VOLUME [CU][7] | AVG TICKET [CU] | RR NO | RR % | VOLUM E [CU] | AVG TICKE T [CU] | RR NO | RR % | VOLUM E [CU] | AVG TICKE T [CU] | RR NO | RR % | VOLUME [CU] | AVG TICKET [CU] |
| CMPG_CP1 | SEG_1 | 11 258 | 2,79% | 240 223 | 21 | 1 913 | 0,47% | 5 409 | 3 | 938 | 0,23% | 1 831 | 2 | 1 723 | 0,43% | 4 125 | 2 |
| CMPG_CP1 | SEG_2 | 2 051 | 2,72% | 36 156 | 18 | 639 | 0,85% | 1 562 | 2 | 654 | 0,87% | 1 618 | 2 | 1 733 | 2,30% | 6 300 | 4 |
| CMPG_CP1 | SEG_3 | 310 | 0,40% | 4 464 | 14 | 58 | 0,08% | 176 | 3 | 18 | 0,02% | 23 | 1 | 42 | 0,06% | 50 | 1 |
| CMPG_CP1 | SEG_4 | 12 227 | 3,76% | 310 662 | 25 | 1 956 | 0,60% | 5 518 | 3 | 1 301 | 0,40% | 2 614 | 2 | 1 791 | 0,55% | 4 068 | 2 |
| CMPG_CP1 | SEG_5 | 390 | 0,32% | 7 590 | 19 | 163 | 0,13% | 635 | 4 | 37 | 0,03% | 77 | 2 | 48 | 0,04% | 245 | 5 |
| *CMPG_CP1_sum* | | *26 236* | *2,61%* | *599 096* | *23* | *4 729* | *0,47%* | *13 299* | *3* | *2 948* | *0,29%* | *6 164* | *2* | *5 337* | *0,53%* | *14 788* | *3* |
| CMPG_CP2 | SEG_1 | 3 794 | 2,59% | 73 176 | 19 | 723 | 0,49% | 1 546 | 2 | 205 | 0,14% | 419 | 2 | 486 | 0,33% | 1 152 | 2 |
| CMPG_CP2 | SEG_2 | 819 | 2,62% | 12 869 | 16 | 250 | 0,80% | 659 | 3 | 144 | 0,46% | 363 | 3 | 559 | 1,79% | 1 703 | 3 |
| CMPG_CP2 | SEG_3 | 101 | 0,42% | 1 640 | 16 | 19 | 0,08% | 53 | 3 | 4 | 0,02% | 7 | 2 | 13 | 0,05% | 15 | 1 |
| CMPG_CP2 | SEG_4 | 2 689 | 4,77% | 63 456 | 24 | 540 | 0,96% | 1 266 | 2 | 177 | 0,31% | 347 | 2 | 426 | 0,75% | 980 | 2 |
| CMPG_CP2 | SEG_5 | 129 | 0,33% | 2 953 | 23 | 41 | 0,10% | 108 | 3 | 12 | 0,03% | 22 | 2 | 14 | 0,03% | 178 | 13 |
| *CMPG_CP2_sum* | | *7 531* | *2,53%* | *154 093* | *20* | *1 573* | *0,53%* | *3 631* | *2* | *542* | *0,18%* | *1 159* | *2* | *1 497* | *0,50%* | *4 029* | *3* |
| CMPG_CP3 | SEG_1 | 2 696 | 1,90% | 53 317 | 20 | 699 | 0,49% | 2 501 | 4 | 452 | 0,32% | 829 | 2 | 583 | 0,41% | 1 435 | 2 |
| CMPG_CP3 | SEG_2 | 591 | 2,58% | 10 850 | 18 | 207 | 0,90% | 510 | 2 | 326 | 1,42% | 860 | 3 | 658 | 2,87% | 3 054 | 5 |
| CMPG_CP3 | SEG_3 | 147 | 0,35% | 1 973 | 13 | 29 | 0,07% | 88 | 3 | 13 | 0,03% | 15 | 1 | 22 | 0,05% | 34 | 2 |
| CMPG_CP3 | SEG_4 | 4 409 | 2,57% | 111 497 | 25 | 833 | 0,49% | 2 637 | 3 | 734 | 0,43% | 1 503 | 2 | 682 | 0,40% | 1 316 | 2 |
| CMPG_CP3 | SEG_5 | 173 | 0,26% | 3 102 | 18 | 102 | 0,15% | 462 | 5 | 16 | 0,02% | 35 | 2 | 26 | 0,04% | 63 | 2 |

[7] CU- Currency Unit

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *CMPG_CP3_sum* | | *8 017* | *1,80%* | *180 739* | *23* | *1 870* | *0,42%* | *6 199* | *3* | *1 540* | *0,35%* | *3 242* | *2* | *1 972* | *0,44%* | *5 903* | *3* |
| CMPG_DP1 | SEG_1 | 403 | 5,34% | 16 148 | 40 | 142 | 1,88% | 618 | 4 | 158 | 2,09% | 495 | 3 | 640 | 8,48% | 12 107 | 19 |
| CMPG_DP1 | SEG_2 | 602 | 1,07% | 42 827 | 71 | 916 | 1,63% | 4 335 | 5 | 893 | 1,59% | 4 727 | 5 | 8 732 | 15,51% | 352 926 | 40 |
| CMPG_DP1 | SEG_3 | 15 | 4,80% | 1 330 | 86 | 11 | 3,54% | 45 | 4 | 4 | 1,22% | 15 | 4 | 38 | 11,80% | 475 | 12 |
| CMPG_DP1 | SEG_4 | 1 413 | 2,94% | 85 953 | 61 | 838 | 1,74% | 3 836 | 5 | 1 012 | 2,10% | 5 174 | 5 | 4 935 | 10,26% | 146 066 | 30 |
| CMPG_DP1 | SEG_5 | 93 | 0,64% | 459 | 5 | 4 | 0,03% | 9 | 2 | 5 | 0,04% | 16 | 3 | 1 886 | 12,89% | 122 | 0,1 |
| *CMPG_DP1_sum* | | *2 526* | *1,99%* | *146 717* | *58* | *1 912* | *1,51%* | *8 843* | *5* | *2 071* | *1,63%* | *10 427* | *5* | *16 231* | *12,79%* | *511 697* | *32* |

After the analysis of the table above there some conclusions were made. The biggest volume of the sales of CP1 and CP2 is in the Campaign with CP1 offer observed. For CP3 and DP1 the biggest volume of sales is noted in Campaign with dedicated DP1 offer. Because of the reason that investigation step includes only Credit Campaign, it occurs that for CP1 the highest response rate is gained in Campaign with CP1 offer. Even if for CP2 the biggest volume is generated in Campaign with CP1 offer, the highest response rate is available in Campaign with CP2. The same dependence is to be seen for the CP3 (the biggest volume in Campaign with CP1 and the highest response rate in Campaign with CP3). It is quite obvious that for DP1 the biggest volume and simultaneously the highest response rate are noted in Campaign with DP1 offer.

In the situation when the customers have been assigned with the clusters from newly proposed MPO segmentation it could be valuable to show the results from a reverse perspective – in the distribution of every Segment.

*Table 5.2.7.6 Summary of Response rates and Volumes achieved in the single-product offering campaigns in division on bank products, regardless of the received offer, from the perspective of proposed segments.*

| | | CP1 | | | | CP2 | | | | CP3 | | | | DP1 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NAME | SEG | RR NO | RR % | VOLUME [CU] | AVG TICKET [CU] | RR NO | RR % | VOLUME [CU] | AVG TICKET [CU] | RR NO | RR % | VOLUME [CU] | AVG TICKET [CU] | RR NO | RR % | VOLUME [CU] | AVG TICKET [CU] |
| CMPG_CP1 | SEG_1 | 11 258 | 2,79% | 240 223 | 21 | 1 913 | 0,47% | 5 409 | 3 | 938 | 0,23% | 1 831 | 2 | 1 723 | 0,43% | 4 125 | 2 |
| CMPG_CP2 | SEG_1 | 3 794 | 2,59% | 73 176 | 19 | 723 | 0,49% | 1 546 | 2 | 205 | 0,14% | 419 | 2 | 486 | 0,33% | 1 152 | 2 |
| CMPG_CP3 | SEG_1 | 2 696 | 1,90% | 53 317 | 20 | 699 | 0,49% | 2 501 | 4 | 452 | 0,32% | 829 | 2 | 583 | 0,41% | 1 435 | 2 |
| CMPG_DP1 | SEG_1 | 403 | 5,34% | 16 148 | 40 | 142 | 1,88% | 618 | 4 | 158 | 2,09% | 495 | 3 | 640 | 8,48% | 12 107 | 19 |
| *SEG_1_sum* | | *18 152* | *2,60%* | *382 864* | *21* | *3 477* | *0,50%* | *10 074* | *3* | *1 752* | *0,25%* | *3 575* | *2* | *3 433* | *0,49%* | *18 819* | *5* |
| CMPG_CP1 | SEG_2 | 2 051 | 2,72% | 36 156 | 18 | 639 | 0,85% | 1 562 | 2 | 654 | 0,87% | 1 618 | 2 | 1 733 | 2,30% | 6 300 | 4 |
| CMPG_CP2 | SEG_2 | 819 | 2,62% | 12 869 | 16 | 250 | 0,80% | 659 | 3 | 144 | 0,46% | 363 | 3 | 559 | 1,79% | 1 703 | 3 |
| CMPG_CP3 | SEG_2 | 591 | 2,58% | 10 850 | 18 | 207 | 0,90% | 510 | 2 | 326 | 1,42% | 860 | 3 | 658 | 2,87% | 3 054 | 5 |
| CMPG_DP1 | SEG_2 | 602 | 1,07% | 42 827 | 71 | 916 | 1,63% | 4 335 | 5 | 893 | 1,59% | 4 727 | 5 | 8 732 | 15,51% | 352 926 | 40 |
| *SEG_2_sum* | | *4 063* | *2,19%* | *102 702* | *25* | *2 012* | *1,08%* | *7 066* | *4* | *2 016* | *1,08%* | *7 568* | *4* | *11 682* | *6,28%* | *363 984* | *31* |
| CMPG_CP1 | SEG_3 | 310 | 0,40% | 4 464 | 14 | 58 | 0,08% | 176 | 3 | 18 | 0,02% | 23 | 1 | 42 | 0,06% | 50 | 1 |
| CMPG_CP2 | SEG_3 | 101 | 0,42% | 1 640 | 16 | 19 | 0,08% | 53 | 3 | 4 | 0,02% | 7 | 2 | 13 | 0,05% | 15 | 1 |
| CMPG_CP3 | SEG_3 | 147 | 0,35% | 1 973 | 13 | 29 | 0,07% | 88 | 3 | 13 | 0,03% | 15 | 1 | 22 | 0,05% | 34 | 2 |
| CMPG_DP1 | SEG_3 | 15 | 4,80% | 1 330 | 86 | 11 | 3,54% | 45 | 4 | 4 | 1,22% | 15 | 4 | 38 | 11,80% | 475 | 12 |
| *SEG_3_sum* | | *573* | *0,40%* | *9 407* | *16* | *118* | *0,08%* | *362* | *3* | *39* | *0,03%* | *61* | *2* | *115* | *0,08%* | *574* | *5* |
| CMPG_CP1 | SEG_4 | 12 227 | 3,76% | 310 662 | 25 | 1 956 | 0,60% | 5 518 | 3 | 1 301 | 0,40% | 2 614 | 2 | 1 791 | 0,55% | 4 068 | 2 |
| CMPG_CP2 | SEG_4 | 2 689 | 4,77% | 63 456 | 24 | 540 | 0,96% | 1 266 | 2 | 177 | 0,31% | 347 | 2 | 426 | 0,75% | 980 | 2 |
| CMPG_CP3 | SEG_4 | 4 409 | 2,57% | 111 497 | 25 | 833 | 0,49% | 2 637 | 3 | 734 | 0,43% | 1 503 | 2 | 682 | 0,40% | 1 316 | 2 |
| CMPG_DP1 | SEG_4 | 1 413 | 2,94% | 85 953 | 61 | 838 | 1,74% | 3 836 | 5 | 1 012 | 2,10% | 5 174 | 5 | 4 935 | 10,26% | 146 066 | 30 |
| *SEG_4_sum* | | *20 737* | *3,45%* | *571 569* | *28* | *4 168* | *0,69%* | *13 256* | *3* | *3 224* | *0,54%* | *9 638* | *3* | *7 833* | *1,30%* | *152 431* | *19* |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CMPG_CP1 | SEG_5 | 390 | 0,32% | 7 590 | 19 | 163 | 0,13% | 635 | 4 | 37 | 0,03% | 77 | 2 | 48 | 0,04% | 245 | 5 |
| CMPG_CP2 | SEG_5 | 129 | 0,33% | 2 953 | 23 | 41 | 0,10% | 108 | 3 | 12 | 0,03% | 22 | 2 | 14 | 0,03% | 178 | 13 |
| CMPG_CP3 | SEG_5 | 173 | 0,26% | 3 102 | 18 | 102 | 0,15% | 462 | 5 | 16 | 0,02% | 35 | 2 | 26 | 0,04% | 63 | 2 |
| CMPG_DP1 | SEG_5 | 93 | 0,64% | 459 | 5 | 4 | 0,03% | 9 | 2 | 5 | 0,04% | 16 | 3 | 1 886 | 12,89% | 122 | 0 |
| *SEG_5_sum* | | *785* | *0,32%* | *14 104* | *18* | *309* | *0,13%* | *1 214* | *4* | *70* | *0,03%* | *151* | *2* | *1 974* | *0,81%* | *608* | *0* |

The table above clearly indicates that in Segment no 1, which focuses on the customers who are the most likely to buy CP1 and CP2, the highest response rates are noted for customers who have got the Deposit offer from the bank. Such a rule is observed for all the proposed products. But while looking at the whole response rates for Segment no 1 and it is worth to stress that the highest response rate is noted for CP1 and then for CP2.

In Segment no 2, which consists of the customers who are likely to buy CP1, CP2 and DP1, the situation is very similar but with one exception for CP1. For CP1 the highest response rate is noted for Campaign with CP1 offer. For the rest of products the best results are achieved in Campaign with DP1 offer. As far as the whole response rates for all purchased products in this Segment are concerned the highest response rate is represented by DP1, then by CP2 and on the same level response rates for remaining CP2 and CP3 are noted.

In the case of Segment no 3, which is described as one where customers are likely to buy DP1, the highest response rates are again the same as in the segments above: for Campaign with DP1 offer. Looking at the summaries of response rates for all the products in this Segment, the best result is noted for CP1, then for DP1 and CP2. But the values of these response rates are much lower than for the two previous Segments and this conclusion is the same as conclusion achived from MPO test.

Distribution of the best response rates for each product in Segment no 4 looks similar to Segment no 2. For CP1 the best outcome is gained for Campaign with CP2 offer. For the remaining three products the best results are placed for Campaign with DP1. To summarize the response rates in this Segments it can be said that the best result was managed by CP1 and DP1. On the third position CP2 is situated.

In the last Segment no 5, which was identified as one with customers who are not likely to buy any product from the bank, for CP1, CP3 and DP1 the best results are achieved in Campaign with DP1 offer. For CP2 the highest response rate is noted in Campaign with CP3 offer. But summing up these results it can be concluded that DP1 gives the best. Sales of CP1 is placed after DP1. On the third place CP2 is placed and sales of CP3 is in the end.

For all the Credit Products the highest response rates, as well as the biggest volume of sales, are available in Segment no 4. For DP1 the best results are to be achieved in Segment no 2. If only customers, who have got the given offer from the bank are taken into account, the situation for three out of four products is the same. To be more precise, for CP1 and CP2 the response rates and volumes of sales are the biggest in Segment no 2. Although for CP3 the best solution is gained by Segment no 4 again, in the set of customers with communication from the bank Segment no 2 presents the highest response rates and the biggest financial effect. In the case of the remaining DP1 and communicated customers Segment no 2 presents the highest outcomes.

Proposed view on the campaign by the prospect of MPO segments allows to track the trends in customers behoviour and allows to define customers who are really likely to buy more than one product. And the most important, it allows to increase the share of CRM actions in total sales of all the bank communication channels.

The cost side left to investigate yet. Values of volume, income, cost and net income are put in the table below.

*Table 5.2.7.7 The financial overview of the single-product campaigns.*

| NAME | OFFER TYPE | SEG | TOTAL VOLUME [CU] | INCOME [CU] | COSTS [CU] | NET INCOME [CU] |
|---|---|---|---|---|---|---|
| CMPG_CP1 | CP1 | SEG_1 | 251 588 | 17 089 | 602 | 16 487 |
| CMPG_CP1 | CP1 | SEG_2 | 45 637 | 2 736 | 113 | 2 623 |
| CMPG_CP1 | CP1 | SEG_3 | 4 713 | 322 | 115 | 207 |
| CMPG_CP1 | CP1 | SEG_4 | 322 862 | 22 011 | 486 | 21 525 |
| CMPG_CP1 | CP1 | SEG_5 | 8 548 | 574 | 185 | 389 |
| *CMPG_CP1_sum* | | | *633 347* | *42 732* | *1 501* | *41 231* |
| CMPG_CP2 | CP2 | SEG_1 | 76 293 | 5 189 | 219 | 4 969 |
| CMPG_CP2 | CP2 | SEG_2 | 15 593 | 964 | 47 | 918 |
| CMPG_CP2 | CP2 | SEG_3 | 1 714 | 117 | 36 | 81 |
| CMPG_CP2 | CP2 | SEG_4 | 66 049 | 4 493 | 84 | 4 409 |
| CMPG_CP2 | CP2 | SEG_5 | 3 261 | 213 | 59 | 154 |
| *CMPG_CP2_sum* | | | *162 912* | *10 977* | *445* | *10 531* |
| CMPG_CP3 | CP3 | SEG_1 | 58 081 | 3 914 | 212 | 3 702 |
| CMPG_CP3 | CP3 | SEG_2 | 15 274 | 854 | 34 | 819 |
| CMPG_CP3 | CP3 | SEG_3 | 2 111 | 143 | 63 | 80 |
| CMPG_CP3 | CP3 | SEG_4 | 116 953 | 7 983 | 256 | 7 727 |
| CMPG_CP3 | CP3 | SEG_5 | 3 662 | 249 | 100 | 149 |
| *CMPG_CP3_sum* | | | *196 082* | *13 143* | *665* | *12 477* |
| CMPG_DP1 | DP1 | SEG_1 | 29 369 | 1 233 | 11 | 1 222 |
| CMPG_DP1 | DP1 | SEG_2 | 404 815 | 4 798 | 84 | 4 714 |
| CMPG_DP1 | DP1 | SEG_3 | 1 865 | 98 | 0 | 97 |
| CMPG_DP1 | DP1 | SEG_4 | 241 030 | 7 056 | 72 | 6 985 |
| CMPG_DP1 | DP1 | SEG_5 | 606 | 34 | 22 | 12 |
| *CMPG_DP1_sum* | | | *677 685* | *13 218* | *190* | *13 029* |

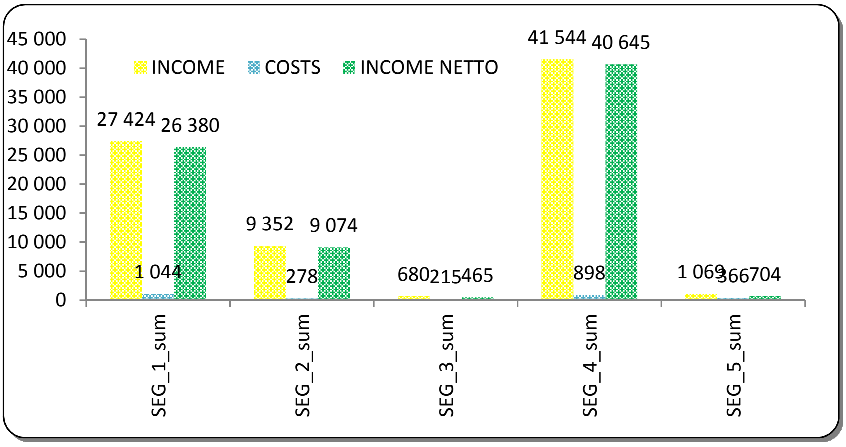
The biggest net profit comes from the Campaign with CP1 offer. Moreover, the costs from this Campaign are also the biggest. The remaining three Campaigns present similar profits. However, the biggest profit is seen in the Campaign with DP1 and the smallest profit is seen in the Campaign with Credit Product no 2.

From the segments perspective results look line in the table below.

*Table 5.2.7.8 The financial overview of the single-product campaigns with MPO segments perspective.*

| NAME | OFFER TYPE | SEG | TOTAL VOLUME [CU] | INCOME [CU] | COSTS [CU] | NET INCOME [CU] |
|---|---|---|---|---|---|---|
| CMPG_CP1 | CP1 | SEG_1 | 251 588 | 17 089 | 602 | 16 487 |
| CMPG_CP2 | CP2 | SEG_1 | 76 293 | 5 189 | 219 | 4 969 |
| CMPG_CP3 | CP3 | SEG_1 | 58 081 | 3 914 | 212 | 3 702 |
| CMPG_DP1 | DP1 | SEG_1 | 29 369 | 1 233 | 11 | 1 222 |
| *SEG_1_sum* | | | *415 332* | *27 424* | *1 044* | *26 380* |
| CMPG_CP1 | CP1 | SEG_2 | 45 637 | 2 736 | 113 | 2 623 |
| CMPG_CP2 | CP2 | SEG_2 | 15 593 | 964 | 47 | 918 |
| CMPG_CP3 | CP3 | SEG_2 | 15 274 | 854 | 34 | 819 |
| CMPG_DP1 | DP1 | SEG_2 | 404 815 | 4 798 | 84 | 4 714 |
| *SEG_2_sum* | | | *481 319* | *9 352* | *278* | *9 074* |
| CMPG_CP1 | CP1 | SEG_3 | 4 713 | 322 | 115 | 207 |
| CMPG_CP2 | CP2 | SEG_3 | 1 714 | 117 | 36 | 81 |
| CMPG_CP3 | CP3 | SEG_3 | 2 111 | 143 | 63 | 80 |
| CMPG_DP1 | DP1 | SEG_3 | 1 865 | 98 | 0 | 97 |
| *SEG_3_sum* | | | *10 404* | *680* | *215* | *465* |
| CMPG_CP1 | CP1 | SEG_4 | 322 862 | 22 011 | 486 | 21 525 |
| CMPG_CP2 | CP2 | SEG_4 | 66 049 | 4 493 | 84 | 4 409 |
| CMPG_CP3 | CP3 | SEG_4 | 116 953 | 7 983 | 256 | 7 727 |
| CMPG_DP1 | DP1 | SEG_4 | 241 030 | 7 056 | 72 | 6 985 |
| *SEG_4_sum* | | | *746 894* | *41 544* | *898* | *40 645* |
| CMPG_CP1 | CP1 | SEG_5 | 8 548 | 574 | 185 | 389 |
| CMPG_CP2 | CP2 | SEG_5 | 3 261 | 213 | 59 | 154 |
| CMPG_CP3 | CP3 | SEG_5 | 3 662 | 249 | 100 | 149 |
| CMPG_DP1 | DP1 | SEG_5 | 606 | 34 | 22 | 12 |
| *SEG_5_sum* | | | *16 077* | *1 069* | *366* | *704* |

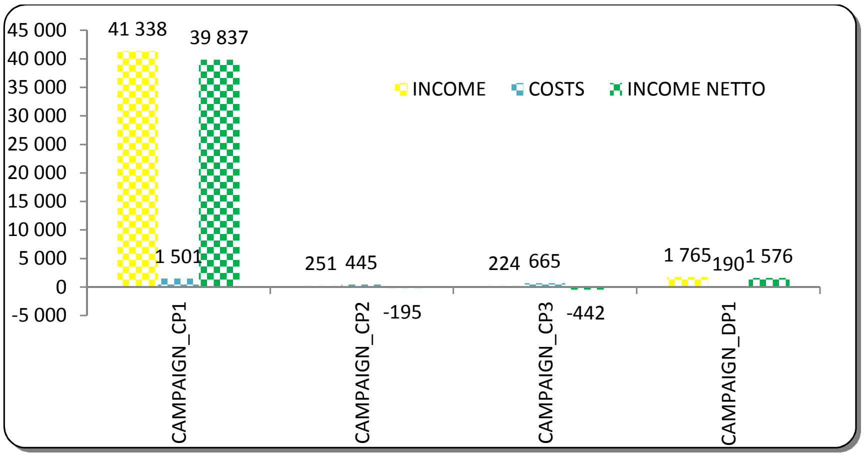
*Figure 5.2.7.1 The financial overview of all the single-product campaigns through assigned multiproduct segments.*




Segment no 4 gives the biggest net profit and comparing it to the second in results – Segment no 1 it produces smaller costs. The last big profit comes from Segment no 2. Segment no 3 and Segment no 5 are the worst as far as net income is concerned.

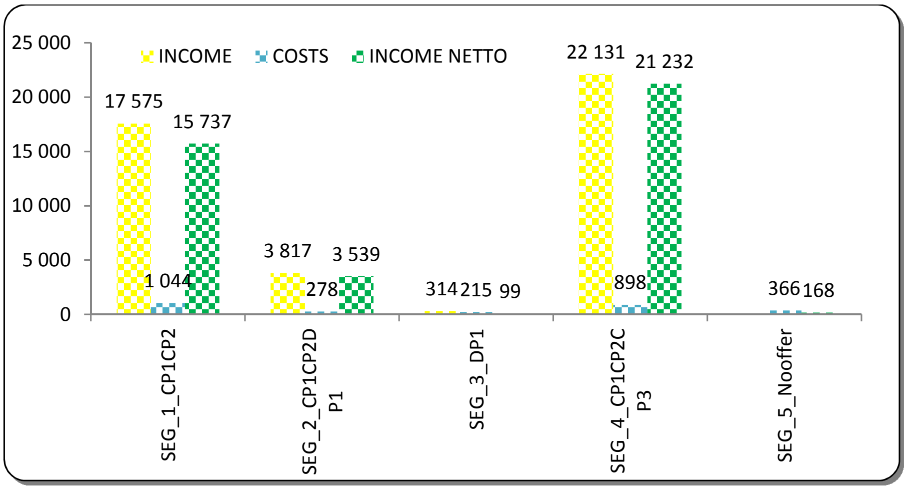
Keeping in mind that the real success of the CRM operations is the action where the customer buys exactly this one product which were proposed in the marketing offer, there is also one more income chart which reveals the profit distribution. This chart presents the profit distribution in particular Campaign (for the first view) and in particular Segments (for the second prospect) and connected only with the communicated product.

*Figure 5.2.7.2 The financial overview of all the single-product campaigns and exactly communicated customers.*



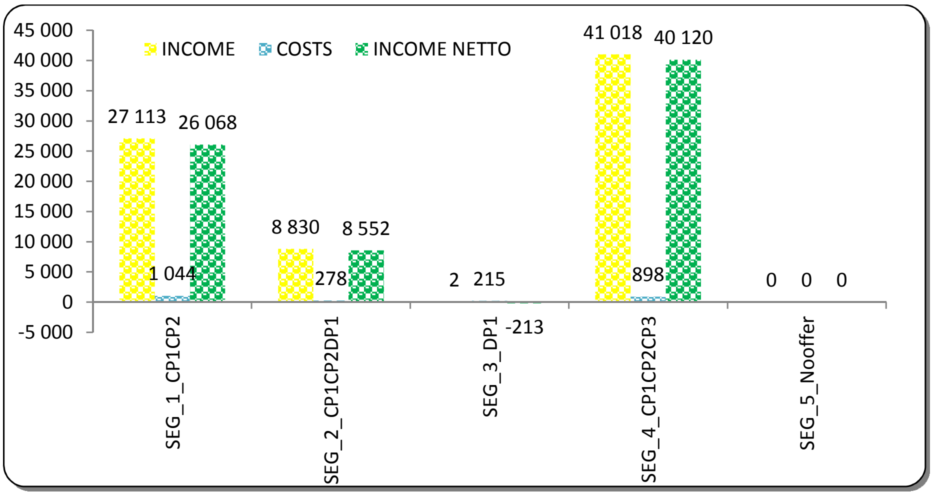

In the case of the really communicated and bought products and recognition of profits Campaign with CP1 is the most profitable. Campaigns with remaining offers of credit products give the income which is not enough to cover the costs. Campaign with DP1 offer brought also the profit, but around twenty-five times less than Campaign with CP1 .

*Figure 5.2.7.3 The financial overview of all the assigned segments and exactly communicated customers.*



Dependence between profits and segments in the groups of communicated customers is the same like in the whole segments. Thus, conclusions for the whole segments are the same as for particular groups described above: the highest profit with almost the same costs is possible to gain in Segment no 4 and in Segment no 1. After these two segments Segment no 2 is placed. Profits of Segment no 3 and Segment no 5 are very low and almost not visible, so as in the MPO test results have showed, there is no sense in offering those two clusters of customers. Besides this summary it is also possible to draw down the potential situation which assumes the main goal of the segmentation. It means that when proposed solution was involved, the customers would be communicated with dedicated group of products instead of only single-product offer as it was used to be for more than one dacade until the MPO approach proposition appeared.

*Figure 5.2.7.4 The financial overview of all the assigned segments and potential situation which use the MPO segments.*




Finally, to **answer the section title question**, which approach gives better results, it is necessary to check if there is a difference between the results of the single-product campaign and theoretical campaign, which could be launched with MPO usage. And if the difference exists – it is possible also to answer how big and what quantitative and volume impact on the results this difference has got. This kind of the comparison of the results could help bank to determine which approach is more profitable.

The table below shows results from the **single-product campaign**. 'Success'here means the purchase, if given product was communicated. To say it comprehensively, if a customer received the offer of the product, which was bought by him/her, the CRM team can say about targeted success which they were able to estimate and predict. But also results from the same single-product offering campaign are displayed, but it shows the **theoretical distribution of the results in the theoretical situation**, if this campaign had been launched using MPO segmentation instead of the single-product offering. As a result each customer was assigned with the MPO segment and 'success' is indicated if customer would have received the communication appropriate to the MPO segment in which the customer would have been in hypothetical comparison. For example, if a customer was typically assigned in Campaign with CP1 offer, he/she received the one offer of CP1. But in the hypothetical situation he/she is also assigned with a given MPO segment, eg. Segment no 4, so he/she would have received the triple offer of CP1, CP2 and CP3. Therefore, purchase of all these three products could be treated as a hypothetical 'success', whereas in the single-product offering only one product (CP1) purchase was a real success. In practice, proposed MPO approach gives the chance to increase the targeted CRM successes just only by putting more products in one letter (according to the propensity-to-buy scores, not the all possible products in one letter).

*Table 5.2.7.9 Summary of response rates and volumes achieved in the single-product offering campaigns in division on the assigned segments from MPO segmentation, according to the appropriate offer in case of theoretical MPO approach.*

| | | SINGLE-PRODUCT CAMPAIGN | | | | HYPOTHETIC MPO TEST IN SINGLE-PRODUCT CAMPAIGN | | | | DIFFERENCES BETWEEN APPROACHES | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NAME | SEG | RR NO | RR % | VOLUME [CU] | AVG TICKET [CU] | RR NO | RR % | VOLUME [CU] | AVG TICKET [CU] | RR NO | RR % | VOLUME [CU] | AVG TICKET [CU] |
| CMPG_CP1 | SEG_1 | 11 258 | 2,79% | 240 223 | 21 | 13 171 | 3,27% | 245 632 | 19 | 1 913 | 0,47% | 5 409 | -3 |
| CMPG_CP2 | SEG_1 | 723 | 0,49% | 1 546 | 2 | 4 517 | 3,08% | 74 722 | 17 | 3 794 | 2,59% | 73 176 | 14 |
| CMPG_CP3 | SEG_1 | 452 | 0,32% | 829 | 2 | 3 395 | 2,40% | 55 818 | 16 | 2 944 | 2,08% | 54 989 | 15 |
| CMPG_DP1 | SEG_1 | 640 | 8,48% | 12 107 | 19 | 545 | 7,22% | 16 767 | 31 | -95 | -1,26% | 4 659 | 12 |
| *SEG_1_sum* | | *13 073* | *1,87%* | *254 705* | *19* | *21 628* | *3,09%* | *392 938* | *18* | *8 555* | *1,22%* | *138 233* | *-1* |
| CMPG_CP1 | SEG_2 | 2 051 | 2,72% | 36 156 | 18 | 4 422 | 5,86% | 44 019 | 10 | 2 372 | 3,14% | 7 862 | -8 |
| CMPG_CP2 | SEG_2 | 250 | 0,80% | 659 | 3 | 1 628 | 5,22% | 15 230 | 9 | 1 378 | 4,42% | 14 572 | 7 |
| CMPG_CP3 | SEG_2 | 326 | 1,42% | 860 | 3 | 1 457 | 6,35% | 14 414 | 10 | 1 131 | 4,93% | 13 555 | 7 |
| CMPG_DP1 | SEG_2 | 8 732 | 15,51% | 352 926 | 40 | 10 249 | 18,21% | 400 088 | 39 | 1 518 | 2,70% | 47 162 | -1 |
| *SEG_2_sum* | | *11 358* | *6,11%* | *390 601* | *34* | *17 756* | *9,55%* | *473 751* | *27* | *6 398* | *3,44%* | *83 150* | *-8* |
| CMPG_CP1 | SEG_3 | 310 | 0,40% | 4 464 | 14 | 42 | 0,06% | 50 | 1 | -268 | -0,35% | -4 414 | -13 |
| CMPG_CP2 | SEG_3 | 19 | 0,08% | 53 | 3 | 13 | 0,05% | 15 | 1 | -6 | -0,03% | -38 | -2 |
| CMPG_CP3 | SEG_3 | 13 | 0,03% | 15 | 1 | 22 | 0,05% | 34 | 2 | 9 | 0,02% | 19 | 0 |
| CMPG_DP1 | SEG_3 | 38 | 11,80% | 475 | 12 | 38 | 11,80% | 475 | 12 | 0 | 0,00% | 0 | 0 |
| *SEG_3_sum* | | *380* | *0,26%* | *5 008* | *13* | *115* | *0,08%* | *574* | *5* | *-265* | *-0,18%* | *-4 434* | *-8* |
| CMPG_CP1 | SEG_4 | 12 227 | 3,76% | 310 662 | 25 | 15 484 | 4,76% | 318 794 | 21 | 3 257 | 1,00% | 8 132 | -5 |
| CMPG_CP2 | SEG_4 | 540 | 0,96% | 1 266 | 2 | 3 406 | 6,04% | 65 069 | 19 | 2 866 | 5,08% | 63 803 | 17 |
| CMPG_CP3 | SEG_4 | 734 | 0,43% | 1 503 | 2 | 5 976 | 3,49% | 115 637 | 19 | 5 242 | 3,06% | 114 134 | 17 |
| CMPG_DP1 | SEG_4 | 4 935 | 10,26% | 146 066 | 30 | 3 263 | 6,78% | 94 963 | 29 | -1 672 | -3,48% | -51 103 | 0 |
| *SEG_4_sum* | | *18 435* | *3,07%* | *459 497* | *25* | *28 128* | *4,68%* | *594 463* | *21* | *9 693* | *1,61%* | *134 966* | *-4* |
| CMPG_CP1 | SEG_5 | 390 | 0,32% | 7 590 | 19 | 0 | 0,00% | 0 | 0 | -390 | -0,32% | -7 590 | -19 |
| CMPG_CP2 | SEG_5 | 41 | 0,10% | 108 | 3 | 0 | 0,00% | 0 | 0 | -41 | -0,10% | -108 | -3 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CMPG_CP3 | SEG_5 | 16 | 0,02% | 35 | 2 | 0 | 0,00% | 0 | 0 | -16 | -0,02% | -35 | -2 |
| CMPG_DP1 | SEG_5 | 1 886 | 12,89% | 122 | 0 | 0 | 0,00% | 0 | 0 | -1 886 | -12,89% | -122 | 0 |
| *SEG_5_sum* | | *2 333* | *0,95%* | *7 855* | *3* | *0* | *0,00%* | *0* | *0* | *-2 333* | *-0,95%* | *-7 855* | *-3* |
| *TOTAL* | | 45 579 | 2,43% | 1 117 666 | 25 | 67 628 | 3,61% | 1 461 727 | 22 | 22 049 | 1,18% | 344 061 | -3 |

This table presents a sum of all purchases for single-product offering campaign and for hypothetical MPO approach. In the last part of the table a difference between those two approaches is showed. The green colour highlights the positive difference and the red colour marks the negative difference. The **biggest positive differences can be observed for Segment no1, Segment no 2 and Segment no 4**. On the other hand, the largest negative differences can be observed in two remaining Segments. Segment no 5 gives the largest negative differences because the MPO segmentation assumes that those customers are not likely to buy any product so they are not offered any product. Also for DP1 the summary above shows that MPO segmentation allows to gain worse results.

While focusing on details – for Segment no 1 changing the single-product campaign into MPO would let **increase response rate** by 65% (from 1.87% to 3.09%) and **increase volume** by 54% (from 255CU[8] to 393CU). Only average ticket of the product would **decrease** by 7% (from 19CU to 18CU). For Segment no 2 – using MPO segmentation would allow to **increase response rates** by 56% (6.11% - 9.55%), **increase volume** by 21% (391CU – 474CU) and **decrease average ticket** by 22% (34CU – 27CU). For Segment no 3 more comfortable would be staying with the single-product offers because proposed MPO approach would **decrease response rates** by 70% (0.26% - 0.08%), **decrease volume** by 89% (5CU – 0.1CU) and average ticket would go down by 62% (13CU –

[8] CU – Currency Unit

5CU). For Segment no 4 response rates with new attitude to offering would **go up** by 53% (3.07% - 4.68%), **volume would increase** by 29% (460CU – 594CU) and average ticket **would decrease** by 15% (25CU – 21CU). For the last Segment (5) MPO would not generate any profits, but it also would not create any costs. In total campaign view MPO approach could **increase response rates** by 48% (from 2.4% to 3.6%), **increase also achieved volumes** by 31% (from 1 118CU to 1 462CU) and unfortunately **decrease average ticke**t by 12% (from 25CU to 22CU).

This summary can be also viewed due to the profitability side. A table with total volume, income, costs and net etto was prepared according to the previous logic lying under the financial campaign overview. This one, which is put below intends to compare the financial overview from the single-product campaign with hypothetical MPO campaign.

*Table 5.2.7.10 The financial overview of the single-product campaign, of the assigned Segments from MPO segmentation and the differences between two approaches.*

| | | ONE-PRODUCT CAMPAIGN | | | | HYPOTHETIC MPO TEST IN SINGLE-PRODUCT CAMPAIGN | | | | DIFFERENCES BETWEEN APPROACHES | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NAME | SEG | TOTAL VOLUME [CU] | INCOME [CU] | COSTS [CU] | NET INCOME [CU] | TOTAL VOLUME [CU] | INCOME [CU] | COSTS [CU] | NET INCOME [CU] | TOTAL VOLUME [CU] | INCOME [CU] | COSTS [CU] | NET INCOME [CU] |
| CMPG_CP1 | SEG_1 | 240 223 | 16 575 | 602 | 15 973 | 245 632 | 16 949 | 602 | 16 346 | 5 409 | 373 | 0 | 373 |
| CMPG_CP2 | SEG_1 | 1 546 | 107 | 219 | -113 | 74 722 | 5 156 | 219 | 4 937 | 73 176 | 5 049 | 0 | 5 049 |
| CMPG_CP3 | SEG_1 | 829 | 57 | 212 | -154 | 55 818 | 3 851 | 212 | 3 640 | 54 989 | 3 794 | 0 | 3 794 |
| CMPG_DP1 | SEG_1 | 12 107 | 42 | 11 | 30 | 16 767 | 1 157 | 11 | 1 146 | 4 659 | 1 115 | 0 | 1 115 |
| *SEG_1_sum* | | *254 705* | *16 781* | *1 044* | *15 737* | *392 938* | *27 113* | *1 044* | *26 068* | *138 233* | *10 332* | *0* | *10 332* |
| CMPG_CP1 | SEG_2 | 36 156 | 2 495 | 113 | 2 382 | 44 019 | 2 624 | 113 | 2 512 | 7 862 | 130 | 0 | 130 |
| CMPG_CP2 | SEG_2 | 659 | 45 | 47 | -1 | 15 230 | 939 | 47 | 893 | 14 572 | 894 | 0 | 894 |
| CMPG_CP3 | SEG_2 | 860 | 59 | 34 | 25 | 14 414 | 794 | 34 | 760 | 13 555 | 735 | 0 | 735 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CMPG_DP1 | SEG_2 | 352 926 | 1 218 | 84 | 1 134 | 400 088 | 4 472 | 84 | 4 388 | 47 162 | 3 254 | 0 | 3 254 |
| *SEG_2_sum* | | *390 601* | *3 817* | *278* | *3 539* | *473 751* | *8 830* | *278* | *8 552* | *83 150* | *5 013* | *0* | *5 013* |
| CMPG_CP1 | SEG_3 | 4 464 | 308 | 115 | 193 | 50 | 0 | 115 | -115 | -4 414 | -308 | 0 | -308 |
| CMPG_CP2 | SEG_3 | 53 | 4 | 36 | -32 | 15 | 0 | 36 | -36 | -38 | -4 | 0 | -4 |
| CMPG_CP3 | SEG_3 | 15 | 1 | 63 | -62 | 34 | 0 | 63 | -63 | 19 | -1 | 0 | -1 |
| CMPG_DP1 | SEG_3 | 475 | 2 | 0 | 1 | 475 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| *SEG_3_sum* | | *5 008* | *314* | *215* | *99* | *574* | *2* | *215* | *-213* | *-4 434* | *-312* | *0* | *-312* |
| CMPG_CP1 | SEG_4 | 310 662 | 21 436 | 486 | 20 949 | 318 794 | 21 997 | 486 | 21 511 | 8 132 | 561 | 0 | 561 |
| CMPG_CP2 | SEG_4 | 1 266 | 87 | 84 | 3 | 65 069 | 4 490 | 84 | 4 405 | 63 803 | 4 402 | 0 | 4 402 |
| CMPG_CP3 | SEG_4 | 1 503 | 104 | 256 | -152 | 115 637 | 7 979 | 256 | 7 723 | 114 134 | 7 875 | 0 | 7 875 |
| CMPG_DP1 | SEG_4 | 146 066 | 504 | 72 | 432 | 94 963 | 6 552 | 72 | 6 481 | -51 103 | 6 049 | 0 | 6 049 |
| *SEG_4_sum* | | *459 497* | *22 131* | *898* | *21 232* | *594 463* | *41 018* | *898* | *40 120* | *134 966* | *18 887* | *0* | *18 887* |
| CMPG_CP1 | SEG_5 | 7 590 | 524 | 185 | 339 | 0 | 0 | 0 | 0 | -7 590 | -524 | -185 | -339 |
| CMPG_CP2 | SEG_5 | 108 | 7 | 59 | -52 | 0 | 0 | 0 | 0 | -108 | -7 | -59 | 52 |
| CMPG_CP3 | SEG_5 | 35 | 2 | 100 | -98 | 0 | 0 | 0 | 0 | -35 | -2 | -100 | 98 |
| CMPG_DP1 | SEG_5 | 122 | 0 | 22 | -21 | 0 | 0 | 0 | 0 | -122 | 0 | -22 | 21 |
| *SEG_5_sum* | | *7 855* | *534* | *366* | *168* | *0* | *0* | *0* | *0* | *-7 855* | *-534* | *-366* | *-168* |
| *TOTAL* | | *1 117 666* | *43 577* | *2 802* | *40 776* | *1 461 727* | *76 962* | *2 436* | *74 527* | *344 061* | *33 385* | *-366* | *33 751* |

In general, the table above consists of the three parts. At this moment it is important to pay more attention on the last one – i.e. the difference between two approaches. The green colour highlights the positive difference and the red colour marks the negative difference. First, looking at the table it is easy to observe that Segment no 1, Segment no 2 and Segment no 4 let achieve **higher net income** in theoretical MPO approach. In Segment no 3 ne **net income is bigger** for the single-product offering. In Segment no 5 in the MPO approach there are **no income and no costs**, because this Segment focuses on the customers who are not likely to buy any product. In the campaign with the single-product offering te customers brought income only if they were proposed CP1. For the remaining products this segment created loss.

Going details, starting with Segment no 1 – the MPO approach would allow to **increase the volume** by 54% (from 255CU to 393CU), **increase the income** by 62% (from 17CU to 27CU). Costs in this Segment are the same because even if MPO segmentation assumes that customer is offered with more than one product (in the case of high probability to buy), physically the offer is put in one letter. That is also a reason why net **income would increase** by 66% (from 16CU to 26CU). The results in Segment no 2 pointed out that MPO approach would allow to **increase the volume** by 21% ( from 391CU to 474CU), but as far as income is concerned there is a growth by 131%. The situation with costs being the same, net income upswing is also very spectacular, it would go up by 142% (from 3,5CU to 8,5CU). In the third segment the MPO test would cause the loss of volume. Volume would decrease by 89% (from 5CU to 0,54CU). As a logically linked chain income would also go down by 99% (from 314CU to 2CU). Costs are the same for both approaches and net income would decrease by 315% (from 99CU to -213CU). As it was previously mentioned, the volume which could be generated by MPO segmentation is 29% higher than in the single-product offering campaign (459CU – 594CU). Income would go up by 85% (from 22CU to 41CU) and costs are the same again, so as a result the MPO approach let increase the net income by 89% (from 21CU to 40CU). For the last Segment the MPO would not generate any profits, but it also would not create any costs. For the single-product offering the net income is quite low – 0,17CU.

In a total campaign view, MPO approach would increase achieved volumes by 31% (from 1 118CU to 1 462CU), would go up with income by 77% (from 44CU to 77 CU), would reduce costs by 13% (from 2,8CU to 2,4CU) and would successfully increase net income by 83% (from 41CU to 75CU).

Even if the table with results in numbers shows that there are a lot of DP1 sold with the high volumes, income and profit from this kind of bank products are not as large and beneficial for finance institutions as credit products are. Credit products let make profits and be more comparative for customers. Therefore, these two results should not be treated the same. Moreover, bank strategy after finance crisis focuses more on an extensible and improving credit side of the bank in order to reduce the financial liquidity, rather than on collecting the deposits and as a result subsidizing the world interest.

**Summarizing this chapter**, it is worth concluding, that MPO approach based on the proposed MPO segmentation allows to detect more numbers of purchases in product distribution and to qualify and count them as the CRM success targeting the right customersm who are really likely to buy the products which are offered. Furthermore, by concluding all available offers in one letter it is easy to **reduce costs** of the campaigns. In the end, when everything is summed up it occurs that such kind of optimization enables to increase all indicators used, from response rates, though volumes and ending with net income and cost reductions.

# 6. Conclusions

In the final chapter, the main conclusions which arose throughout this research are drawn, practical implementations are clarified and opportunities for further research are specified and highlighted.

## 6.1 Main conclusions

The main objective of this doctoral thesis was to present a new approach to provide customers with a retail bank product offer. The solution developed is the opposite of what has been used in the financial institution so far. To be more precise, the solution is to change the current single-product offering to MPO approach, which in practice means that a bank customer does not receive the letter with a single-product offer with the highest probability to buy (very often the same product every several months), but receives offer consists of those bank products, for which customer probability to buy score is the highest. The estimated scores are determined by the propensity-to-buy econometric models created for all bank products.

In the first of two major parts of the thesis, which consists of three chapters, the initial business background for the single-product offering campaign in the cross-sell process has been shown. It is one of the key elements of this part of the research, because every action of analytical CRM team is preceded by a defined business need to share. In this case, the need consists of determining the likelihood of buying four key banking products for every customer of the bank. The assignment of such a propensity score allows to rank all the customers starting from the most likely to the least inclined to buy the bank product. Therefore, created customers hierarchy can be divided into ten groups forming the so-called. deciles of probability. This allows to move from a continuous variable, which in this case is the probability of purchase, for a discrete variable, which creates a set of 10 values, ranging from 1 to 10, where 1 is the group of customers most likely to buy a bank product and defines the range of maximum likelihood of the willingness to buy bank product.

Using the available variables for each of the four key products, there were propensity-to-buy those particular product models developed. The propensity-to-buy model is designed to determine the likelihood of the mentioned purchase of a particular product for a specific customer. The model is created on the basis of historical purchasing events, and also defines the propensity-to-buy a particular product by a particular customer, in particular opportunities that in a simplest terms means the given purchase of given product after receiving offers from the bank. In this section everyone can find a whole methodology to create such a model, the results of the various stages of modeling and indicative values for given response rates on banking offer. Easy to register is a rule that for a product more popular among customers model resets better results at each stage of the modeling process than for the products which are bought less willingly. Moreover, propensity-to-buy models usage approach highlights the advantage this method in comparison with experts` knowledge usage, because models allow to rank customers according to the propensity-to-buy scores as well as accoridin to the estimated response rates.

The last section of this part describes the assumptions and logic behind the proposed analytical MPO segmentation approach. It explains how to successfully connect customers in five homogeneous groups using their propensity-to-buy scores which are defined for four basic bank products.

Such a summary can find the relations between created scores and it also allows to gain the awareness that customers who are typical for this bank form some groups. It occurs that in this specific finance institution there are customers who are interested in buying CP1, CP2 and DP1 (25% of the customer database), customers interested in buying CP1, CP2 (11%) and one more segment full of debtors – customers interested in CP1, CP2, CP3 (10%). Two remaining segments cover more than 54% of database. They focus on customers who are interested in DP1 (the biggest group, 31%) and on customers who are in general not interested in any bank product. Those customers have the lowest scores in all the created propensity-to-buy models.

Newly proposed clustering solution can also increase the feeling that customer can be treated as a full unit, not only as a customer who is interested in one product. Such an approach can be more accurate with the CRM ideology, which assumes that customer should be placed in the middle of all fields of interest, what leads to changing the organization from the product-centric into customer-centric. Furthermore, by linkning all scores in one center, the planning and conducting some steps towards optimization methods process seem to look easier and more confident.

The second part of the thesis focuses on a new campaign offering process, in which the customer receives various proposals of the purchase opportunities of thebank products from the bank. These proposals depends on the segment to which the customer was able to be assigned according to created MPO segmentation. MPO segmentation is based on existing deciles of four propensity-to-buy models. In order to determine if the client behaves in the way that was found by a segment description, it was necessity to create a campaign based on the assumptions of MPO segmentation approach. The purpose and details of the MPO segmentation approach test are precisely described in the second part of the study. Also in this section the obtained results are presented, along with the full comparisons of the test groups and the control groups, on which the main conclusions have been drawn. The first conclusion claims that **customers with Segment no 3** (customers likely to buy DP1) **and Segment no 5** (customers not likely to buy any bank product) **should not be offered any more because the response rates are very low**, there are no business and even no analytical justifications for low indicators and there are no significant differences between test and control groups. It was counted that pausing the offering customers from these two segments would allow to save 1 004CU. The volume which could be generated by offered customers could be 740CU, so it shows that loss could be more than 260CU. From the marketing perspective those saved 1 004CU could be used for more profitable campaigns or marketing promotions instead of wasting for offering customers with low propensity-to-buy scores.

Secondly, as far as Segment no 1 is concerned the results between test and control groups show that the customers are really likely to buy two Credit Products: CP1 and CP2 and for this type of customers these **two bids** should be combined.

Thirdly, the biggest response rates and the highest results are presented by Segment no 4, which consist of the typical consumer finance products likely to attract customers. The hypothesis for this group has been confirmed, these **customers should be offered with a combination of products offer for CP1, CP2 and CP3** since customers with a triple offer create high response rates for all the Credit Products.

The fourth conclusion proves the effectiveness of the CRM work by showing the results that **customers who are getting the offer in the campaign are more likely to purchase bank products than those, who have not received any mailing**. The average difference between customers who have received the offer and those without ranges from 46% (in Segment no 1) to 51% (in Segment no 4).

The fifth conclusion concerns Segment no 2. Test showed that customers who formed this group have a high awareness and strong financial need and **they are coming to buy bank products regardless of the offer**

**received.** It means that they really do not need the offer from the bank because they are very interested in buying bank products. Therefore, this is the potential field of costs reduction. Although, this is the first hypothesis, which is recommended to be verified and checked with more campaigns.

The penumltimate chapter presents the different recognitions and comparisons of two approaches and highlights the dominant position of the newly proposed solutions. It also emphasizes the **advantage of MPO segmentation approach in increasing the basic indicators and rates which are key measures determining the degree of success of each campaign**. The superiority of MPO segmentation approach lies in **increasing the total response rate** from the campaign by 48%, increasing the total generated volume by 31%, increasing the total income by 77% and total net income by 83%. The dominance is also stressed by the recognition that costs can be reduced by 13% (by not offering the customers who come from the last Segment no 5). The new approach is a kind of the answer to the optimization huge database problem. The results show that this solution meets the basic manager`s need since the increase is very high, both in the response rates and also in profit from the campaign. Proposed MPO segmentation approach reveals that changing the long-term process can change the way of looking at and perceiving the customers and it would also benefit to recogniz the potential of new ways to target, offer, and find savings and to indicate the results and successes. To put it more simply – change the product-centric organization into customer-centric organization.

The role of data mining models in marketing is quite new. Although expanding rapidly, data mining is still 'foreign territory' for many marketers who trust only their 'intuition' and domain experience. Their segmentation schemes and marketing campaign lists are created by the business rules based on their business knowledge. Data mining models are not 'threatening': they cannot substitute or replace the significant role of the domain experts and their business knowledge. These models, although powerful, cannot effectively work without the active support of the business experts. On the contrary, only when data mining capabilities are complemented with business expertise, they can achieve truly meaningful results. For instance, the predictive ability of a data mining models can be substantially increased by including informative inputs with predictive power suggested by experienced persons in the field. Additionally, the information of existing business rules/scores can be integrated into a data mining model and contribute to the building of a more robust and successful result. Moreover, before the actual deployment model results should always be evaluated by business experts with respect to their meaning, in order to minimize the risk of coming up with trivial or unclear findings. Thus, business domain knowledge can truly help and enrich the data mining results. On the other hand, data mining models can identify patterns that even the most experienced business people may have missed. They can help in fine tuning the existing business rules, and enrich, automate, and standardize judgmental ways of working which are based on the personal perceptions and views. They comprise an objective, data-driven approach, minimizing subjective decisions and simplify time-consuming processes. In conclusion, the combination of business domain expertise with the power of data mining models can help organizations gain a competitive advantage in their efforts to optimize customer management.

## 6.2 Further research

Finally, the thesis highlights some opportunities for further research. First, further research can take into account different value propositions connected with specific product and given segment. As customers from all segments are different across the segments, but very similar inside the cluster, they would probably be interested in different attributes of products between segments but very similar inside the group.

A key challenge could be developing propensity-to-buy models for all the existing products and then creating a segmentation based on all scores. It is possible that within the segments more subsegments could be found. It could open more doors to analyses, results and interpretations and then develop and launch the strategy which could be specified for given segment or even subsegment.

The proposed in this research MPO could also be a one of the several attributes describing the bank customers. Based on selected differentiated variables the general bank customers` segmentation could be developed. It would describe and define customers who formed specified groups. If during verification and validation results were rewarding and satisfactory it could be a good premise and motivator to trsnsform the overall bank strategy from being product oriented to customer segments oriented. Then, if profitability could be added to the segmentation as a one of the dimensions it would help managers to determine segments of customers. From the business point of view these customers (or rather all segments) which bank would be able to earn the most on them but also these customers who are the least profitable for the institution, would be the most interested to distinguish. Finally, profitability is correlated with the customer value. This customer value could be well defined and then combined with propensity-to-buy scores. In the end the customer life cycle could be formulated, what would additionally confirm and justify the business suitability of the proposed framework.

# Appendices

## A. Methodology details

### A.1 Logistic Regression and Stepwise Selection formulas

Fisher scoring algorithm is connected with its mathematical form, which is presented below.

Let`s consider that $\boldsymbol{Z_{ij}}=(Z_{1j},\ldots,Z_{k+1,j})'$ is multinomial variable:

$$Z_{ij} = \begin{cases} 1 & if\ Y_j = i \\ 0 & otherwise \end{cases} \tag{A.1.1}$$

$P_{ij}$ denotes the probability that the $j$ th observation has response value $i$ and the expected value of $\boldsymbol{Z_j}$ is

$$P_j=(P_{1j},\ \ldots,\ P_{k+1,j})' \tag{A.1.2}$$

$$\text{while } P_{k+1,j} = 1 - \sum_{i=1}^{k} P_{ij}. \tag{A.1.3}$$

The mentioned covariance matrix of $\boldsymbol{Z_j}$ is $\boldsymbol{V_j}$ which in practice is the covariance matrix of a multinomial random variable for one trial with parameter vector $P_j$.

Let $\beta$ be the vector of regression parameters

$$\beta = (\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_s)', \tag{A.1.4}$$

and let $\boldsymbol{D_j}$ be the matrix of partial derivatives of $P_j$ with the respect to β. The estimating equation for the regression parameters is

$$\sum_j D_j' W_j \left(Z_j - P_j\right) = 0, \tag{A.1.5}$$

where

$$W_j = w_j f_j V_j', \tag{A.1.6}$$

and $w_j$ is the weight $f_j$ is the frequency of the $j$ th observation, and $\boldsymbol{V}_j'$ is a generalized inverse of $\boldsymbol{V}_j$ and it is the inverse of the diagonal matrix with $P_j$ as the diagonal.

The maximum likelihood estimate of $\beta$ is obtained iteratively with starting point of value of $\beta(0)$ as:

$$\beta(m+1) = \beta(m) + \left(\sum_i D_j' W_j D_j\right)^{-1} \sum_i D_j' W_j (Z_j - P_j), \tag{A.1.7}$$

where $\boldsymbol{D}_j$, $\boldsymbol{W}_j$ *and* $P_j$ are evaluated at $\beta(m)$. The step size is the expression after the plus sign. If the likelihood evaluated at next $\beta(m+1)$ is less than that which was counted at $\beta(m)$, then $\beta(m+1)$ is recomputed by step-halving. The iterative scheme continues till convergence is gained – what means, till $\beta(m+1)$ is sufficiently close to $\beta(m)$. Then the maximum likelihood estimated of $\beta$ is:

$$\hat{\beta} = \beta(m+1). \tag{A.1.8}$$

The covariance matrix of $\hat{\beta}$ is estimated in turn by formula:

$$\widehat{Cov}(\widehat{\beta}) = \left(\sum_j \widehat{D}_j' \widehat{W}_j \widehat{D}_j\right)^{-1} \tag{A.1.9}$$

where $\widehat{D}_j$ and $\widehat{W}_j$ are $\boldsymbol{D}_j$ and $\boldsymbol{W}_j$ evaluated at $\hat{\beta}$. For the intercept parameters starting values are the observed cumulative logits, what means logits of the observed cumulative proportions of response.

Stepwise selection option is similar to the forward selection in the beginning. First parameters for effects forced into the model are estimated. These effects are the intercepts and the first *n* explanatory effects in the model statement. Next, the score chi-square statistic for each effect not in the model is computed and the largest of these statistics is examines. If it is significant at the slentry=level, the corresponding effect is added to the model. Once an effect is entered in the model, it does not necessarily remain. It is different from forward in this point as there effects which are entered into a model already, they are never removed from the model. In stepwise selection effects are entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. This process terminates if no further effect can be added to the model or if the current model is identical to a previously visited model. The significance values that were used in this project, they were also modified from default with following values: slentry=0.07 (0.05 by default) – is required to allow a variable into the model and slstay=0.5 (0.7 by default) – is required for a variable to stay in the model.

After estimated the probability of the observed response ($\widehat{P_j}$) for the $j$th observation, in order to set the model fitness to data, three criteria are calculated:

I. -2 Log Likelihood:

$$-2\ Log\ L = -2\ \sum_j \frac{w_j}{\sigma^2} f_j \log(\widehat{P_j}) \qquad \text{(A.I.10)}$$

where $w_j$ is the weight value and $f_j$ is the frequency value of the $j$ th observation and $\sigma^2$ is the dispersion parameter, which equals 1. For binary response models that arre present in every built model, where events and trials as the modeled variable are used, this criterion is equivalent to:

$$-2\ Log\ L = -2\ \sum_j \frac{w_j}{\sigma^2} f_j [r_j \log(\widehat{P_j}) + (n_j - r_j) \log(1 - \widehat{P_j})] \qquad \text{(A.I.11)}$$

where $r_j$ is the number of events, $n_j$ is the number of trials and $\widehat{P_j}$ is the estimated event probability.

II. Akaike Information Criterion

$$AIC = -2\ Log\ L + 2p \qquad \text{(A.I.12)}$$

where $p$ is the number of parameters in the model. For cumulative response models,

$$p = k + s \qquad \text{(A.I.13)}$$

where $k$ is the total number of response levels minus one and $s$ is the number of explanatory effects. For the generalizes model

$$p = k\ (s + 1) \qquad \text{(A.I.14)}$$

III. Schwarz (Bayesian Information) Criterion:

$$SC = -2\ Log\ L + p \log(\textstyle\sum_j f_j) \quad \text{(A.I.15)}$$

where $p$ is the number of parameters in the model.

The *AIC* and *SC* statistics give two different ways of adjusting the (-2 Log L) statistic for the number of terms in the model and the number of observations used. These statistics can be used when comparing different models for the same data. Lower values of the statistics indicate a more desirable model.

# B. Scoring step details

## B.1 Propensity-to-buy Credit Product no 1 model - details

This section focuses on the detailed results of the created propensity-to-buy Credit Product no1 model. It is a kind of a supplement of subchapter 4.2.1.

Table no. B.1.1 lists **the parameter estimates**, their standard errors and the results of the **Wald test** for individual parameters. It is shown below:

*Table B.1.1 Analysis of Maximum Likelihood Estimates for CP1.*

| Analysis of Maximum Likelihood Estimates | | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Parameter type | Parameter description | DF | Estimator | Standard Error | Wald Chi-Square | Pr > Chi-Sq. |
| CP1_v1_d4 | account-related dummy | customer has got given products in rare group of products | 1 | -0.5710 | 0.0263 | 472.7306 | <.0001 |
| CP1_v2_d4 | account-related dummy | sum of customer debts is more than 3000 PLN | 1 | 0.1250 | 0.0122 | 105.3033 | <.0001 |
| CP1_v3_d1 | account-related dummy | the original interest rate of the first product | 1 | 0.2183 | 0.0606 | 12.9664 | 0.0003 |
| CP1_v3_d2 | account-related dummy | the original interest rate of the first product is >0 and <6.3 | 1 | 0.4813 | 0.0622 | 59.7939 | <.0001 |
| CP1_v3_d3 | account-related dummy | the original interest rate of the first product is >6.3 and <10.3 | 1 | 0.4409 | 0.0632 | 48.5893 | <.0001 |
| CP1_v3_d4 | account-related dummy | the original interest rate of the first product is >10.3 | 1 | 0.5939 | 0.0605 | 96.5318 | <.0001 |
| CP1_v4_d1 | account-related dummy | risk grade is <0.07 | 1 | -0.7777 | 0.0620 | 157.5922 | <.0001 |
| CP1_v4_d2 | account-related dummy | risk grade is >0.07 and <0.13 | 1 | -0.7956 | 0.0575 | 191.4005 | <.0001 |
| CP1_v4_d3 | account-related dummy | risk grade is >0.13 and <0.19 | 1 | -0.7319 | 0.0553 | 175.1783 | <.0001 |
| CP1_v4_d4 | account-related dummy | risk grade is >0.19 and <0.25 | 1 | -0.7488 | 0.0543 | 189.9504 | <.0001 |
| CP1_v4_d5 | account-related dummy | risk grade is >0.25 and <0.29 | 1 | -0.7969 | 0.0563 | 200.0711 | <.0001 |
| CP1_v4_d6 | account-related dummy | risk grade is >0.29 and <0.38 | 1 | -0.6623 | 0.0522 | 160.8275 | <.0001 |
| CP1_v4_d7 | account-related dummy | risk grade is >0.38 | 1 | -0.4688 | 0.0497 | 89.0083 | <.0001 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CP1_v5_d1 | account-related dummy | customer does not have any credit products reported in credit bureau | 1 | -0.1071 | 0.0201 | 28.4375 | <.0001 |
| CP1_v5_d2 | account-related dummy | customer does not have any credit products with previous main bank | 1 | -0.1971 | 0.0145 | 184.4518 | <.0001 |
| CP1_v5_d3 | account-related dummy | customer does not have any credit products no 2 reported in credit bureau | 1 | -0.0826 | 0.0190 | 18.8595 | <.0001 |
| CP1_v6_d1 | account-related dummy | customer has got no delays in paying off the loan | 1 | 0.2120 | 0.0259 | 66.8213 | <.0001 |
| CP1_v6_d2 | account-related dummy | customer has got delay in pating off the loan (1-19 days) | 1 | 0.3623 | 0.0350 | 107.3517 | <.0001 |
| CP1_v7_d1 | account-related dummy | maximum days of delay in paying off the loan =0 | 1 | -0.2500 | 0.0145 | 295.7657 | <.0001 |
| CP1_v7_d2 | account-related dummy | maximum days of delay in paying off the loan is 1-2 | 1 | -0.1590 | 0.0197 | 65.3675 | <.0001 |
| CP1_v7_d3 | account-related dummy | maximum days of delay in paying off the loan is 3-4 | 1 | -0.0966 | 0.0185 | 27.2214 | <.0001 |
| CP1_v8_d23 | account-related dummy | customer deposit balance is <1500 PLN | 1 | 0.3261 | 0.0173 | 355.3584 | <.0001 |
| CP1_v8_d36 | account-related dummy | customer deposit balance is >10000 PLN | 1 | -0.3888 | 0.0344 | 127.6857 | <.0001 |
| CP1_v9_d43 | account-related dummy | customer credit balance is < 1000 PLN | 1 | 0.3009 | 0.0251 | 143.6128 | <.0001 |
| CP1_v10 | CRM-related | customer was communicated via phone in last year | 1 | -0.0281 | 0.00970 | 8.3679 | 0.0038 |
| CP1_v11 | CRM-related | customer was communicated in given season | 1 | 0.3221 | 0.0121 | 713.3877 | <.0001 |
| CP1_v12_d1 | CRM-related dummy | customer has not bought credit product no 1 in bank campaign | 1 | -0.2652 | 0.0252 | 110.9895 | <.0001 |
| CP1_v13_d8 | CRM-related dummy | customer has bought any product in bank campaign | 1 | 0.1722 | 0.0195 | 77.9882 | <.0001 |
| CP1_v14_d1 | demographic dummy | customer income <635 PLN | 1 | -0.2688 | 0.0461 | 34.0092 | <.0001 |
| CP1_v14_d3 | demographic dummy | customer income >1200 PLN | 1 | 0.0619 | 0.0168 | 13.5300 | 0.0002 |
| CP1_v15_d3 | demographic dummy | customer is single or married | 1 | -0.1991 | 0.0148 | 181.1338 | <.0001 |
| CP1_v16_d1 | demographic dummy | customer graduated given education no 0 | 1 | -0.6892 | 0.0217 | 1008.4750 | <.0001 |
| CP1_v16_d2 | demographic dummy | customer graduated given education no 1 | 1 | 0.5633 | 0.0147 | 1467.3523 | <.0001 |
| CP1_v17_d1 | demographic dummy | any information about customer`s children | 1 | -0.8192 | 0.0139 | 3492.6686 | <.0001 |
| CP1_v17_d2 | demographic dummy | customer has got 1 child | 1 | 0.0796 | 0.0149 | 28.3696 | <.0001 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CP1_v18_d2 | demographic dummy | customer has got given agreement type | 1 | -0.1761 | 0.0339 | 27.0462 | <.0001 |
| CP1_v18_d5 | demographic dummy | customer has got given agreement type | 1 | -0.3522 | 0.0135 | 683.9007 | <.0001 |
| CP1_v19_d5 | demographic dummy | customer lives in given district | 1 | 0.0491 | 0.0149 | 10.8633 | 0.0010 |
| CP1_v20_d1 | demographic dummy | customer works in given occupation | 1 | -0.2431 | 0.0180 | 181.7397 | <.0001 |
| CP1_v20_d3 | demographic dummy | customer works in given occupation | 1 | -0.2697 | 0.0124 | 475.2728 | <.0001 |
| CP1_v21_d1 | demographic dummy | customer has got internet bank service access | 1 | 0.2234 | 0.0134 | 279.4350 | <.0001 |
| CP1_v22_d1 | demographic dummy | customer has got no dependent persons | 1 | -0.2373 | 0.0301 | 62.3141 | <.0001 |
| CP1_v23_d1 | demographic dummy | credit risk bureau rate =0 | 1 | -0.3833 | 0.0136 | 799.6817 | <.0001 |
| CP1_v23_d4 | demographic dummy | credit risk bureau rate >3 and <5 | 1 | 0.0793 | 0.0166 | 22.8299 | <.0001 |
| CP1_v24_d3 | demographic dummy | customer lives in given district | 1 | -0.1294 | 0.0230 | 31.6863 | <.0001 |
| CP1_v25_d3 | demographic dummy | customer lives in given district | 1 | -0.6261 | 0.0464 | 181.8928 | <.0001 |
| CP1_v25_d4 | demographic dummy | customer lives in given district | 1 | -0.4423 | 0.0236 | 350.1179 | <.0001 |
| CP1_v25_d5 | demographic dummy | customer lives in given district | 1 | -0.3599 | 0.0233 | 238.0945 | <.0001 |
| CP1_v25_d6 | demographic dummy | customer lives in given district | 1 | -0.2892 | 0.0438 | 43.5588 | <.0001 |
| CP1_v25_d7 | demographic dummy | customer lives in given district | 1 | -0.2301 | 0.0335 | 47.1299 | <.0001 |
| CP1_v25_d8 | demographic dummy | customer lives in given district | 1 | -0.2319 | 0.0185 | 157.0850 | <.0001 |
| CP1_v25_d9 | demographic dummy | customer lives in given district | 1 | -0.1280 | 0.0159 | 64.8422 | <.0001 |
| CP1_v25_d10 | demographic dummy | customer lives in given district | 1 | -0.0621 | 0.0150 | 17.2065 | <.0001 |
| CP1_v25_d15 | demographic dummy | customer lives in given district | 1 | 0.1271 | 0.0189 | 45.1728 | <.0001 |
| CP1_v25_d16 | demographic dummy | customer lives in given district | 1 | 0.1603 | 0.0262 | 37.5410 | <.0001 |
| CP1_v25_d17 | demographic dummy | customer lives in given district | 1 | 0.2468 | 0.0199 | 153.4385 | <.0001 |
| CP1_v26_d5 | demographic dummy | customer has given zip code of his/her place of living | 1 | -0.1134 | 0.0160 | 49.9637 | <.0001 |
| CP1_v26_d6 | demographic dummy | customer has given zip code of his/her place of living | 1 | -0.0776 | 0.0147 | 27.7325 | <.0001 |
| CP1_v26_d7 | demographic dummy | customer has given zip code of his/her place of living | 1 | -0.1439 | 0.0247 | 33.8779 | <.0001 |
| CP1_v26_d8 | demographic dummy | customer has given zip code of his/her place of living | 1 | -0.0276 | 0.0133 | 4.2819 | 0.0385 |
| CP1_v26_d11 | demographic dummy | customer has given zip code of his/her place of living | 1 | 0.1307 | 0.0333 | 15.4271 | <.0001 |
| CP1_v26_d1 | demographic dummy | customer has given zip code of his/her place of living | 1 | 0.3023 | 0.0110 | 751.1599 | <.0001 |
| CP1_v27_d11 | demographic dummy | customer is at 18-27 age | 1 | 0.1689 | 0.0182 | 86.0897 | <.0001 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CP1_v28_d52 | demographic dummy | customer new income >2500 PLN and <3500 PLN | 1 | 0.0757 | 0.0133 | 32.6250 | <.0001 |
| CP1_v28_d53 | demographic dummy | customer new income >3500 PLN and <4500 PLN | 1 | 0.1213 | 0.0193 | 39.5788 | <.0001 |
| CP1_v28_d54 | demographic dummy | customer new income >4500 PLN | 1 | 0.1348 | 0.0169 | 63.9727 | <.0001 |
| CP1_v29_d55 | demographic dummy | months of customer bank history <6 months | 1 | 0.8987 | 0.0494 | 331.3183 | <.0001 |
| CP1_v29_d56 | demographic dummy | months of customer bank history 6-12 months | 1 | 0.5892 | 0.0294 | 402.4955 | <.0001 |
| CP1_v29_d57 | demographic dummy | months of customer bank history 12-24 months | 1 | 0.4202 | 0.0208 | 407.0435 | <.0001 |
| CP1_v29_d58 | demographic dummy | months of customer bank history 24-36 months | 1 | 0.3996 | 0.0193 | 427.8094 | <.0001 |
| CP1_v29_d59 | demographic dummy | months of customer bank history 36-48 months | 1 | 0.2358 | 0.0195 | 145.9214 | <.0001 |
| CP1_v29_d60 | demographic dummy | months of customer bank history 48-60 months | 1 | 0.1530 | 0.0188 | 65.9326 | <.0001 |
| CP1_v29_d61 | demographic dummy | months of customer bank history 60-72 months | 1 | 0.0941 | 0.0184 | 26.1659 | <.0001 |
| CP1_v29_d65 | demographic dummy | months of customer bank history >108 months | 1 | -0.0687 | 0.0143 | 23.1732 | <.0001 |
| CP1_v30_d69 | demographic dummy | months since last bought product >24 and <36 months | 1 | -0.4642 | 0.0154 | 910.0250 | <.0001 |
| CP1_v30_d70 | demographic dummy | months since last bought product >36 and <48 months | 1 | -0.8232 | 0.0206 | 1591.3636 | <.0001 |
| CP1_v30_d71 | demographic dummy | months since last bought product >48 and <60 months | 1 | -0.9868 | 0.0247 | 1596.6464 | <.0001 |
| CP1_v30_d72 | demographic dummy | months since last bought product >60 and <72 months | 1 | -1.0330 | 0.0332 | 969.8702 | <.0001 |
| CP1_v30_d73 | demographic dummy | months since last bought product >72 and <84 months | 1 | -0.8970 | 0.0491 | 333.3628 | <.0001 |
| CP1_v30_d74 | demographic dummy | months since last bought product >84 months | 1 | -0.7027 | 0.0551 | 162.8109 | <.0001 |
| CP1_v31 | products characteristics | customer has got given credit group of products no 1 | 1 | -0.4978 | 0.0260 | 367.7463 | <.0001 |
| CP1_v32 | products characteristics | customer has got given credit product no 1 | 1 | 0.8543 | 0.0177 | 2335.4209 | <.0001 |
| CP1_v33 | products characteristics | customer has got given deposit product no 2 | 1 | 0.1411 | 0.0291 | 23.5428 | <.0001 |
| CP1_v34 | products characteristics | customer has got given credit product no 4 | 1 | -0.1787 | 0.0152 | 137.5606 | <.0001 |
| CP1_v35_d5 | products characteristics | customer has got given credit group of products no 2 | 1 | -0.8425 | 0.0224 | 1412.7995 | <.0001 |
| CP1_v36_d7 | products characteristics | customer has got given deposit product no 1 | 1 | -0.6147 | 0.0338 | 329.9187 | <.0001 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CP1_v37_d28 | products characteristics | customer has got given credit product no 5 | 1 | 0.1639 | 0.0580 | 7.9970 | 0.0047 |
| CP1_v38_d32 | products characteristics | customer has got given credit product no 6 | 1 | 0.1215 | 0.0378 | 10.3227 | 0.0013 |
| CP1_v39_d33 | products characteristics | customer has got given deposit group of products no 1 | 1 | 0.0748 | 0.0318 | 5.5254 | 0.0187 |
| CP1_v40_d36 | products characteristics | customer has got given credit group of products no 3 | 1 | 0.2662 | 0.0183 | 211.4471 | <.0001 |
| CP1_v41_d1 | products characteristics dummy | sum of credit group of products no 1 =1 | 1 | 0.2000 | 0.0172 | 135.8309 | <.0001 |
| CP1_v41_d6 | products characteristics dummy | sum of credit products no 1 is >2 but <6 | 1 | 0.1156 | 0.0132 | 76.8676 | <.0001 |
| CP1_v42_d1 | products characteristics dummy | sum of credit group of products no 4 =1 | 1 | -0.1204 | 0.0108 | 124.0282 | <.0001 |
| CP1_v42_d3 | products characteristics dummy | sum of credit group of products no 4 >3 | 1 | 0.1562 | 0.0211 | 54.6707 | <.0001 |
| CP1_v43_d2 | products characteristics dummy | sum of credit product no 2 =2 | 1 | 0.1471 | 0.0191 | 59.4434 | <.0001 |
| CP1_v44_d2 | products characteristics dummy | customer has got 2 different products | 1 | 0.2400 | 0.0236 | 103.4118 | <.0001 |
| CP1_v44_d3 | products characteristics dummy | customer has got 3 different products | 1 | 0.3217 | 0.0235 | 187.5507 | <.0001 |
| CP1_v44_d4 | products characteristics dummy | customer has got 4 different products | 1 | 0.5367 | 0.0244 | 483.9564 | <.0001 |
| CP1_v44_d5 | products characteristics dummy | customer has got 5-10 different products | 1 | 0.7110 | 0.0238 | 894.2105 | <.0001 |
| CP1_v44_d6 | products characteristics dummy | customer has got 10-42 different products | 1 | 0.7998 | 0.0278 | 824.8847 | <.0001 |
| CP1_v45_d1 | sociodemographic | customer gives salary into bank account | 1 | 0.6377 | 0.0165 | 1488.4120 | <.0001 |
| CP1_v46 | sociodemographic | customer is new | 1 | 0.1649 | 0.0216 | 57.9973 | <.0001 |
| CP1_v0 | Intercept | Intercept | 1 | -2.5261 | 0.0937 | 726.3327 | <.0001 |

Variables ending with ‘d’ means that this variable is dummy variable of the original one. Moreover, variables CP1_v14-30 and CP1_v44-v45 are customer-related and they give sociodemographic picture of customer, variables CP1_v24-v26 describe how geographical characteristics influences on buying the Credit Product no1 by customer. Variables CP1_v10-v13 are linked with CRM campaigns history and frequency and way of contact with customer. Variables CP1_v30-v44 say about the history of customer portfolio and its size in monetary aspect. Variables CP1v1-v13 are connected with customers' accounts and they inform about the payment history and customer tendency to being on time with his/her commitments. Although for several variables probability of **chi-square statistics** were>0.0001 while reducing these variables made the model worse predicting, since they are still present in the model. From a statistical point of view linear hypothesis for $\beta$estimators are expressed in matrix form as

$$H_0: L\beta = c \tag{B.1.1}$$

where $\boldsymbol{L}$ is a matrix of coefficients for the linear hypothesis, and *c* is a vector of constants. The vector of regression coefficients $\boldsymbol{\beta}$ includes slope parameters as well as intercepts parameters. The Wald chi-square statistic for testing mentioned $H_0$ is computed as follows:

$$\lambda_W^2 = \left(\boldsymbol{L}\hat{\beta} - c\right)' [\boldsymbol{L}\widehat{\boldsymbol{V}}(\hat{\beta})\boldsymbol{L}']^{-1}(\boldsymbol{L}\hat{\beta} - c) \tag{B.1.2}$$

where $\widehat{\boldsymbol{V}}\,(\hat{\beta})$ is the estimated covariance matrix. $\lambda_W^2$ has an asymptotic chi-square distribution with *r* degrees of freedom and it is the rank of $\boldsymbol{L}$.

**The Association of Predicted Probabilities and Observed Responses** results, which are showed by Table B.1.2 display measures of the association between predicted probabilities and observed responses which include a breakdown of the number of pairs with different responses and four rank correlation indexes. After launched the model on Development sample of CP1 there were 81% of Concordant pairs, meaning for 81% of 1,3492E+11 pairs model estimated correctly higher probability for higher ordered value than for lower ordered value. For 16,5% of total number of pairs model gave higher probability for lower ordered value and for 2,5% pairs probability for both ordered values was the same.

*Table B.1.2 Association of Predicted Probabilities and Observed Responses for CP1.*

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Concordant | 81.0 | Somers' D | 0.645 |
| Percent Discordant | 16.5 | Gamma | 0.661 |
| Percent Tied | 2.5 | Tau-a | 0.029 |
| Pairs | 1,3492E+11 | c | 0.823 |

Mentioned above four measures of association for assessing the predictive ability of a model are based on the number of pairs of observations with different response values, the number of concordant pairs and the number of discordant pairs. A pair of observation with different observed responses is said to be concordant if the observation with the lower ordered response value has a lower predicted mean score than the observation with the higher ordered response value. If the observation with the lower ordered response value has a higher predicted mean score than the observation with the higher ordered response value, then the pair is discordant. If the pair is neither concordant nor discordant, it is a tie. The predicted mean score of an observation is the sum of the Ordered Values, shown by Table 4.2.1.1 (subchapter 4.2.1) Response Profile, minus one, weighted by the corresponding predicted probabilities for that observation:

$$\text{P}redicted\ mean\ score = \sum_i^{k+1}(i-1)\widehat{P_i} \quad \text{(B.1.3)}$$

where *(k+1)* is the number of response levels and $\widehat{P_i}$ is the predicted probability of *i* th (ordered) response.

Enumeration of the total number of concordant and discordant pairs is carried out by categorizing the predicted mean score into intervals of length k=500 and accumulating the corresponding frequencies of observations. Let *N* be the sum of observation frequencies in the data. Suppose there are a total of *t* pairs with different responses: $n_c$ of them are concordant and $n_d$ of them are discordant and $(t - n_c - n_d)$ are tied. The following four indices or rank correlation for assessing the predictive ability of a model look as below:

$$c = \frac{(n_c + 0.5(t - n_c - n_d)}{t} \quad \text{(B.1.4)}$$

$$Somers`\ D\ Gini\ coefficient = \frac{(n_c - n_d)}{t} \tag{B.1.5}$$

$$Goodman - Kruskal\ Gamma = \frac{(n_c - n_d)}{(n_c + n_d)} \tag{B.1.6}$$

$$Kendall`s\ Tau - a = \frac{(n_c - n_d)}{(0.5N(N-1))} \tag{B.1.7}$$

If there is no ties then **Somers` D** (Gini coefficient) = *(2c-1)*. It is also worth to emphasize that index *c* also gives an estimate of the area under the receiver operating characteristics (ROC) curve when, as in describing case – response is binary.

The **Hosmer-Lemeshow** goodness-of-fit statistics, which results are to find in subchapter 4.2.1 are obtained by calculating the Pearson chi-square statistic from the *2 x g* table of observed and expected frequencies, where *g* is the number of groups. The statistic is written:

$$\lambda_{HL}^2 = \sum_{i=1}^{g} \frac{(O_i - N_i \bar{P_i})^2}{N_i \bar{P_i}(1 - \bar{P_i})} \tag{B.1.8}$$

where $N_i$ is the total frequency of subject in the *i* th group, $O_i$ is the total frequency of event outcomes in the *i* th group and $\bar{P_i}$ is the average estimated predicted probability of an event outcome for the *i* th group. The Hosmer-Lemeshow statistic is compared to a chi-square distribution with *(g-n)* degrees of freedom, where the value of *n* is 2 by default, but it is possible to change. Large values of $\lambda_{HL}^2$ and small p=values indicate a lack of fit of the model.

Getting back to $\bar{P_i}$ variable it will be necessary to present how predicted and confidence limits are calculated by using the maximum likelihood estimated (MLEs). For a vector of explanatory variables *X*, the linear predictor is presented in a form of equation:

$$\eta_i = g(\Pr(Y \leq i|x)) = \alpha_i + x'\beta \quad for\ \ 1 \leq i \leq k \tag{B.1.9}$$

and it is estimated as follows:

$$\hat{\eta_i} = \hat{\alpha_i} + x'\hat{\beta} \tag{B.1.10}$$

where $\hat{\alpha_i}$ and $\hat{\beta}$ are the maximum likelihood estimators of $\alpha_i$ and $\beta_i$ parameters.

The estimated standard error of $\eta_i$ is $\hat{\sigma}(\hat{\eta}_i)$ and it can be calculated as the square root of the quadratic form:

$$(1, x')\widehat{\boldsymbol{V}}_{\boldsymbol{b}}(1, x')' \tag{B.1.11}$$

where $\widehat{\boldsymbol{V}_{\boldsymbol{b}}}$ is the estimated covariance matrix of parameter estimated. The asymptotic 100(1-ɑ)% confidence interval for $\eta_i$ is given with formula:

$$\widehat{\eta_i} \pm z_{\alpha/2}\hat{\sigma}(\widehat{\eta_i}) \tag{B.1.12}$$

where $z_{\alpha/2}$ is the 100(1-ɑ/2) percentile of a standard normal distribution. The predicted probability for:

$$P_i = \Pr(Y \leq i|x) \tag{B.1.13}$$

is get by back-transforming the corresponding measures for the linear predictor and as a result for **Logit link function** it is captured by formula:

$$\widehat{P_i} = \frac{1}{(1+\exp(-\widehat{\eta_i}))} \tag{B.1.14}$$

which is synonymus with the general formula presented by (4.2.2). In the same conditions the 100(1-ɑ)% confidence limits are defined as follows:

$$\frac{1}{(1+\exp\left(-\widehat{\eta_i} \pm z_{\alpha/2}\hat{\sigma}(\widehat{\eta_i})\right))} \tag{B.1.15}$$

For binary response data, the response is either an Event or a Nonevent (1 or 0). The response with Ordered Value 1 is regarded as the Event and the response with Ordered Value 2 (Response Profile Table) is the Nonevent. Logistic Regression modeling process models the probability of the Ordered Value 1, so of the Event. From the fitted model, a predicted Event probability can be computed for each observation. If the predicted Event probability exceeds or equals some cutpoint value $z \in [0,1]$, (usually equals 0.5) the observation is predicted to be an Event observation. Otherwise, it is predicted as a Nonevent. A 2 × 2 frequency table can be obtained by cross-classifying the observed and predicted responses.

The **accuracy** of the classification is measured by its sensitivity - the ability to predict an Event correctly, and specificity - the ability to predict a Nonevent correctly. **Sensitivity** is the proportion of Event responses that were predicted to be the Event. Specificity is the proportion of Nonevent responses that were predicted to be Nonevent. There is also possible to count three other condition probabilities: false positive rate, false negative rate and rate of correct classification. The first one, false positive rate is the proportion of predicted Event responses that were observed as Nonevents. The second rate, the false negative rate is the proportion of predicted Nonevent responses that were observed as Events. Table B.1.3 shows **Classification table**.

*Table B.1.3 Classification table results of propensity-to-buy CP1 model.*

| Classification table | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Prob | Correct | | Incorrect | | Percentage | | | | | |
| Level | Event | Nonevent | Event | Nonevent | Correct | Sensitivity | Specificity | False POS | False NEG | |
| 0.500 | 31682 | 975418 | 7935 | 218484 | 79.9 | 81.7 | 83.5 | 20.1 | 18.3 | |

In order to present this measures in more statistical form lets` assume that in a sample of $n$ individuals $n_1$ individuals are observed to have a certain event. This group is denoted by $C_1$. The group of remaining $n_2=n-n_1$ individuals who do have nonevent is denoted by $C_2$. Risk factors are identified for the sample and a logistic regression is fitted to the data. For the *j*th individual, by using MLE, an estimated probability $\widehat{P_i}$ of the event is calculated. Then, lets` suppose that $n$ individuals undergo a test for predicting the event and the test is based on the estimated probability of the event. Higher values of the estimated probability are assumed to be associated with the events. The following formulas show definitions of measures which are presented in Table B.1.4.

$$POS(z) = \sum_{i\in C_1} I(\widehat{P_i} \geq z) \tag{B.1.16}$$

$$NEG(z) = \sum_{i\in C_2} I(\widehat{P_i} < z) \tag{B.1.17}$$

$$False\ POS(z) = \sum_{i\in C_2} I(\widehat{P_i} \geq z) \tag{B.1.18}$$

$$False\ NEG(z) = \sum_{i\in C_1} I(\widehat{P_i} < z) \tag{B.1.19}$$

$$SENS(z) = \frac{POS(z)}{n_1} \tag{B.1.20}$$

$$1SPEC(z) = \frac{False\ POS(z)}{n_2} \tag{B.1.21}$$

where *I(.)* is the indicator function. *POS(z)* is the number of correctly predicted event responses, *NEG(z)* is the number of correctly predicted nonevent responses, *False POS(z)* is the number of falsely predicted event responses, *False NEG(z)* is the number of falsely predicted nonevent responses. *SENS(z)* is the sensitivity of the test and *1SPEC(z)* is one minus specificity of the test. According to the table above a developed model has represented correctness of 79.9% of observations on the 0.5 probability level. When target variable is 1 there are 31 682 observations which are corrected assigned and for target variable equals 0 there are 975 418 correct assignments. Moreover, for 20.1% of observations model gave the 0 value while it should be 1 value and for the little less percentage, 18.3% it gave the 1 value while it should be 0 value.

The last step of modeling process is Model validation. The Development set is used for learning, i.e. fitting the model parameters And Validation set is used to tune these parameters and to minimize overfitting the model, what means that model correctness should be checked on the Validation sample.

Below Figure B.1.1 shows the ROC curves for Development and Validation sample of propensity-to-buy CP1 model. It occurs that created model fits the observations and it is going to predict marketing offer responders correctly.

*Figure B.1.1 ROC charts for development and validation data sets of propensity-to-buy CP1 model.*

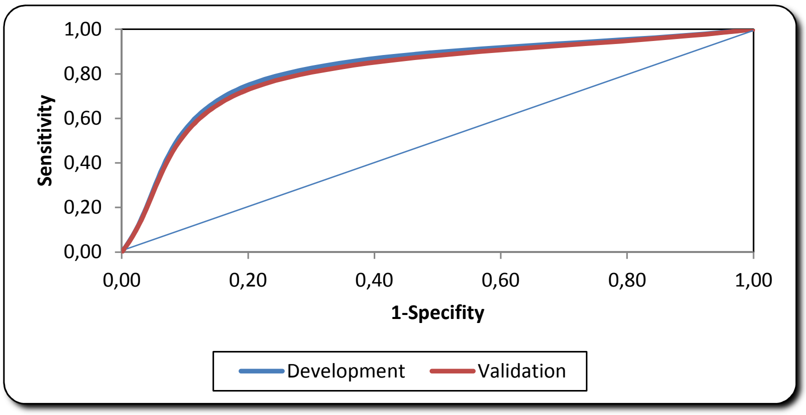


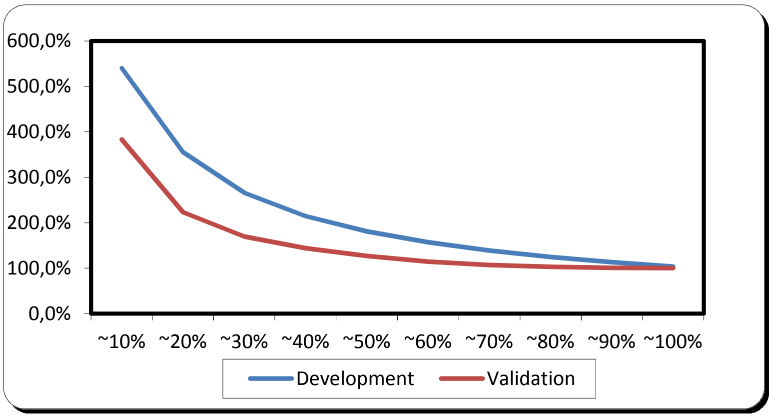

Table B.1.4 shows the number of Events and Nonevents and **Lift rate** value for each decile after applying the propensity-to-buy for CP1 model on the Validation data set.

*Table B.1.4 Events and Nonevents results and Lift rate for application propensity-to-buy CP1 model on the validation sample.*

| Validation sample | | | | | | |
|---|---|---|---|---|---|---|
| Decile | Non Resp. | Resp. | Total | Resp. Rate | Cum. Resp rate | % Lift |
| ~10% | 123 210 | 16 619 | 139 829 | 11,89% | 11,89% | 382,9% |
| ~20% | 153 124 | 3 913 | 157 037 | 2,49% | 6,92% | 222,8% |
| ~30% | 126 689 | 1 877 | 128 566 | 1,46% | 5,27% | 169,7% |
| ~40% | 100 014 | 1 122 | 101 136 | 1,11% | 4,47% | 144,0% |
| ~50% | 90 797 | 796 | 91 592 | 0,87% | 3,94% | 126,8% |
| ~60% | 83 097 | 564 | 83 661 | 0,67% | 3,55% | 114,3% |
| ~70% | 58 554 | 338 | 58 892 | 0,57% | 3,32% | 106,9% |
| ~80% | 33 686 | 161 | 33 847 | 0,48% | 3,20% | 103,0% |
| ~90% | 17 889 | 77 | 17 966 | 0,43% | 3,13% | 101,0% |
| ~100% | 8 877 | 29 | 8 906 | 0,32% | 3,10% | 100,0% |
| | 795 934 | 25 496 | 821 430 | 3,10% | | |

While applying the model on the Validation sample, the cutpoints for each decile are rewritten from the Development sample. That is the reason why the deciles in Validation sample have got the different frequencies then the standard rule that decile is 10% of the data set. Although, the rule is still the same, the highest responders are put into decile 1 and the lowest into 10 decile. Comparing the results of Validation sample to those of the Development sample (subchapter 4.2.1) it concludes that they are very similar. In the top decile there were 129 829 customers, with 16 619 responders and 11.89% of response rate. As average response rate is 3.10%, Lift rate equals 3.83 (383%) what is worse result than in Development sample. Figure below shows the comparison of Lift rates for Development and Validation data sets.

*Figure B.1.2 Lift ratio results charts for development and validation data sets of CP1 model.*

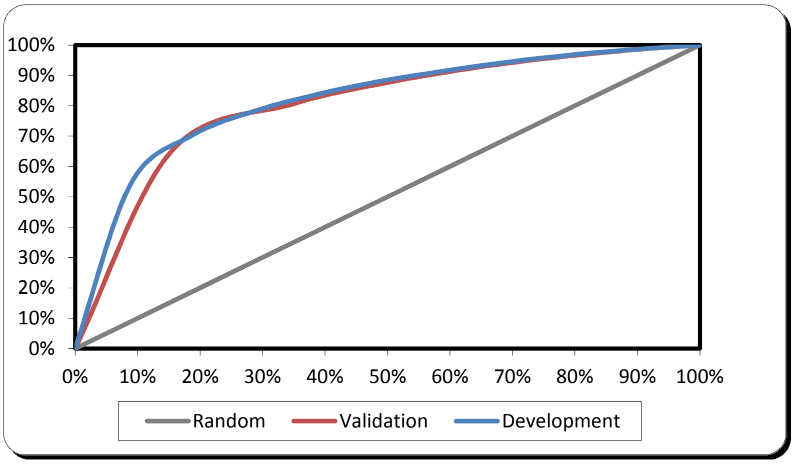



Another chart compares the cumulative percent of responses captured as each decile is added to the target. According to the table B.1.5 prepared for Validation data set the top two deciles capture 80.5% of responders what means that in this case the rule '80/20' is met.

*Table B.1.5 Event and Nonevent distribution in deciles for validation sample of propensity-to-buy for CP1 model.*

| Model Validation all APPL | | | | | | | |
|---|---|---|---|---|---|---|---|
| Approx. score % | NonResp (NR) | Responders (R) | Prob. NonResp | % of all NonResp | % of all Resp | Cum. % NonResp | Cum. % Resp |
| ~10% | 123 210 | 16 619 | 0,881 | 15,5% | 65,2% | 15,5% | 65,2% |
| ~20% | 153 124 | 3 913 | 0,975 | 19,2% | 15,3% | 34,7% | 80,5% |
| ~30% | 126 689 | 1 877 | 0,985 | 15,9% | 7,4% | 50,6% | 87,9% |
| ~40% | 100 014 | 1 122 | 0,989 | 12,6% | 4,4% | 63,2% | 92,3% |
| ~50% | 90 797 | 796 | 0,991 | 11,4% | 3,1% | 74,6% | 95,4% |
| ~60% | 83 097 | 564 | 0,993 | 10,4% | 2,2% | 85,0% | 97,6% |
| ~70% | 58 554 | 338 | 0,994 | 7,4% | 1,3% | 92,4% | 99,0% |
| ~80% | 33 686 | 161 | 0,995 | 4,2% | 0,6% | 96,6% | 99,6% |
| ~90% | 17 889 | 77 | 0,996 | 2,2% | 0,3% | 98,9% | 99,9% |
| ~100% | 8 877 | 29 | 0,997 | 1,1% | 0,1% | 100,0% | 100,0% |
| Totals | 795 934 | 25 496 | | | | | |

*Figure B.1.3 Comparison of cumulative percent of Events and Nonevents for development and validation data set of CP1 model.*




Model represents enough quality of goodness-of-fit between the set of explanatory variables and the target variable. Both curves are above the random baseline.

## B.2 Propensity-to-buy Credit Product no 2 model - details

This model was built in the second place and it is also based on the historical data of previous actions regarding the behavior of customers in similar situations. Its goal is to predict how much given customer is interested in buying CP2 within the marketing action. As in the previous case, but this time connected with CP2, there were many marketing campaigns which took place the whole year before the time of building the model, since campaigns had to happen regularly. As a result it was like nine main campaigns with mailing contact and nine followers campaigns (additional campaign to the same customers but with the other form of contact) with phone contact and seven other, additional campaigns with mailing and phone contact in the input data set. Observation window, in the numbers included 2 002 451 customers who were targeted, 8 109 who filled in the credit application and 5 614 who really bought the CP2. It gives response rate equals 0.40% for customers with filled in application and 0.28% for real buyers. The same rule about target variable was used here: people who filled in the credit application become the target variable=1 means 'good'. The rest, who was offered and did not use it became target variable=0 means 'bad. Response Profile table is presented below:

*Table B.2.1 Response Profile for CP2.*

Response Profile

| Ordered Value | CP1 | Total Frequency |
|---|---|---|
| 1 | Event \| 1 | 8 109 |
| 2 | NonEvent \| 0 | 1 994 342 |

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied

Since the response rates in this case were very small, the way of creating development and validation samples was different this time, just to adapt to the results of the marketing campaigns. There were ten development samples created, which consisted on the whole available Events and the same number of NonEvents. NonEvents were random selected from the whole NonEvent data set. In this approach development sample has got an equal number of both values of target variable. Based on the created development samples ten propensity-to-buy models for CP2 were built. Finally, after the in-depth analyses of each developed model the best one was created. Afterwards as a validation sample the remaining nine samples were used and the whole large data set the model was also applied on.

Willingness of purchasing CP2 can be described with the following formula:

$$P_{CP2} = Prob(Y_{CP2}|X_{CP2}) = \frac{1}{1+e^{-L_{CP2}}} \quad \text{(B.2.1)}$$

where

$$L_{CP2} = \alpha + \beta_1 X_{1CP2} + \beta_2 X_{2CP2} + \dots + \beta_k X_{kCP2} \quad \text{(B.2.2)}$$

where *CP2* means Credit Product no 2. Below there are placed all general and detailed results which propensity-to-buy for CP2 model was described.

*Table B.2.2 Analysis of Maximum Likelihood Estimates for CP2.*

| Analysis of Maximum Likelihood Estimates | | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Parameter type | Parameter description | DF | Estimator | Standard Error | Wald Chi-Square | Pr > Chi-Sq. |
| CP2_v1 | account-related | credit risk bureau rate | 1 | 0.00183 | 0.000143 | 164.8325 | <.0001 |
| CP2_v2 | CRM-related | customer has bought any product in bank campaign | 1 | 0.9757 | 0.3076 | 10.0621 | 0.0015 |
| CP2_v3 | CRM-related | customer was communicated via phone in last 6 months | 1 | 0.5105 | 0.0802 | 40.5114 | <.0001 |
| CP2_v4 | CRM-related | customer was communicated via mail in last 6 months | 1 | 0.8314 | 0.0824 | 101.8064 | <.0001 |
| CP2_v5 | CRM-related | customer has bought credit product no 1 in bank campaign | 1 | 2.9355 | 1.0681 | 7.5543 | 0.0060 |
| CP2_v6_d2 | demographic dummy | months of customer bank history 28-56 months | 1 | 0.3499 | 0.0914 | 14.6664 | 0.0001 |
| CP2_v7_d1 | demographic dummy | customer income 0 PLN and <1237 PLN | 1 | -0.4081 | 0.1545 | 6.9755 | 0.0083 |
| CP2_v7_d2 | demographic dummy | customer new income >1237 PLN and <1795 PLN | 1 | 0.6719 | 0.0988 | 46.2880 | <.0001 |
| CP2_v7_d4 | demographic dummy | customer new income >3000 PLN and <3749 PLN | 1 | 0.6824 | 0.1313 | 27.0087 | <.0001 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CP2_v8 | products characteristics | customer has got deposit product no 3 | 1 | 1.9002 | 0.0904 | 442.3060 | <.0001 |
| CP2_v9 | products characteristics | customer has got credit product no 1 | 1 | 1.3325 | 0.0784 | 289.0578 | <.0001 |
| CP2_v10 | products characteristics | sum of credit group of products no 2 | 1 | 0.7138 | 0.0475 | 225.4633 | <.0001 |
| CP2_v11 | products characteristics | customer has got deposit product no 2 | 1 | 1.1399 | 0.1315 | 75.1299 | <.0001 |
| CP2_v12 | products characteristics | customer has got credit product no 5 | 1 | -0.4769 | 0.0808 | 34.8497 | <.0001 |
| CP2_v13 | products characteristics | customer has got credit product no 4 | 1 | -1.3698 | 0.1123 | 148.9024 | <.0001 |
| CP2_v0 | | | 1 | -3.0913 | 0.1115 | 769.3257 | <.0001 |

All the estimated parameters are significantly different from zero ($p<0.05$). It simply means that all used variables could be interpreted and they well explain the estimated customer propensity-to-buy for CP2 in the marketing campaign. Variables ending with 'd' means that this variable is dummy variable of the original one. Variables CP2_v6-v8 are customer-related and they describe customer from demographic point of view mostly, variables CP2_v2-v5 are linked with CRM campaigns history and frequency and way of contact with customer. Variables CP2_v8-v13 say about the history of customer portfolio and CP2_v1 is connected with customer risk history.

The Association of Predicted Probabilities and Observed Responses results, which are showed by the Table B.2.3 display measures of the association between predicted probabilities and observed responses which include a breakdown of the number of pairs with different responses and four rank correlation indexes. After launched the model on the development sample of CP2 there were 90.6% of Concordant pairs, meaning for almost 91% of 1,01E+07 pairs model estimated correctly higher probability for higher ordered value than for lower ordered value. For 16,5% of total number of pairs model gave higher probability for lower ordered value and for 2,5% pairs probability for both ordered values was the same.

*Table B.2.3 Association of Predicted Probabilities and Observed Responses for CP2.*

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Concordant | 90.6 | Somers' D | 0.814 |
| Percent Discordant | 9.2 | Gamma | 0.815 |
| Percent Tied | 0.2 | Tau-a | 0.407 |
| Pairs | 10144226 | c | 0.907 |

Odds ratios estimates for the parameters used in propensity-to-buy for CP2 model are able to find below.

*Table B.2.4 Odds ratio estimates for CP2.*

| Odds Ratio Estimates | | | | |
|---|---|---|---|---|
| Effect | Parameter description | Point Estimate | 95% Wald Confidence Limits | |
| CP2_v1 | credit risk bureau rate | 1.002 | 1.002 | 1.002 |
| CP2_v10 | sum of credit group of products no 2 | 2.042 | 1.860 | 2.241 |
| CP2_v11 | customer has got deposit product no 2 | 3.126 | 2.416 | 4.046 |
| CP2_v11 | customer has got deposit product no 2 | 0.621 | 0.530 | 0.727 |
| CP2_v12 | customer has got credit product no 5 | 0.254 | 0.204 | 0.317 |
| CP2_v2 | customer has bought any product in bank campaign | 2.653 | 1.452 | 4.848 |
| CP2_v3 | customer was communicated via phone in last 6 months | 1.666 | 1.424 | 1.950 |
| CP2_v4 | customer was communicated via mail in last 6 months | 2.296 | 1.954 | 2.699 |
| CP2_v5 | customer has bought credit product no 1 in bank campaign | 18.832 | 2.321 | 152.763 |
| CP2_v6_d2 | months of customer bank history 28-56 months | 1.419 | 1.186 | 1.697 |
| CP2_v7_d1 | customer income >0 PLN and <1237 PLN | 0.665 | 0.491 | 0.900 |
| CP2_v7_d2 | customer new income >1237 PLN and <1795 PLN | 1.958 | 1.613 | 2.376 |
| CP2_v7_d4 | customer new income >3000 PLN and <3749 PLN | 1.979 | 1.530 | 2.559 |
| CP2_v8 | customer has got deposit product no 3 | 6.687 | 5.602 | 7.982 |
| CP2_v9 | customer has got credit product no 1 | 3.790 | 3.251 | 4.420 |

In general, the odds ratio shows the strength of the association between a predictor and the response of interest. It can vary from 0 to infinity. In the propensity-to-buy CP2 model there is one variable for which odds ratio is almost equal 1. It means that such variable has got no association with the target variable The remaining variables are greater than 1 (11 variables) or smaller than 1 (3 variables). For variable CP2_v3, where odds ratio=1.666 it can be interpreted as follows: if there are customers who are different from each other only in terms of being communicated via phone in last 6 months, the customer, who was communicated via phone in last 6 months has an odds ratio of buying CP2 versus not buying CP2 which is 66,6% higher than for the customers who was not communicated. For variable CP2_v7_d1, where odds ratio=0.665 it can be in turn described in a way: if there are customers who are different from each other only in terms of income, which is smaller than 1237 PLN, the customer who has so small income, between 0-1237 PLN has an odds ratio of buying CP2 versus not buying CP2 which is about 43,5% lower than the customer whose income equals 0 OLN.

The table B.2.5 illustrates the Hosmer and Lemeshow test results of propensity-to-buy CP2 model.

*Table B.2.5 The Hosmer-Lemeshow goodness-of-fit results of propensity-to-buy CP2 model.*

| Hosmer and Lemeshow Goodness-of-Fit Test | | |
|---|---|---|
| Chi-square | DF | Pr>Chi-square |
| 4,545 | 8 | 0,3246 |

| Hosmer-Lemeshow test results | | | | | |
|---|---|---|---|---|---|
| Group | Total | Target_variable=1 | | Target_variable=0 | |
| | | Observed | Predicted | Observed | Predicted |
| 1 | 205 904 | 11 | 7 | 205 893 | 205 897 |
| 2 | 194 901 | 18 | 15 | 194 883 | 194 886 |
| 3 | 200 245 | 39 | 30 | 200 206 | 200 215 |
| 4 | 200 245 | 77 | 72 | 200 168 | 200 173 |
| 5 | 202 760 | 114 | 109 | 202 646 | 202 651 |
| 6 | 198 988 | 150 | 154 | 198 838 | 198 834 |
| 7 | 200 245 | 273 | 276 | 199 972 | 199 969 |
| 8 | 200 874 | 471 | 480 | 200 403 | 200 394 |
| 9 | 200 245 | 1 109 | 1 112 | 199 137 | 199 133 |
| 10 | 198 045 | 5 849 | 5 854 | 192 195 | 192 191 |

Table B.2.6 displays the Classification results of created propensity-to-buy for CP2 model.

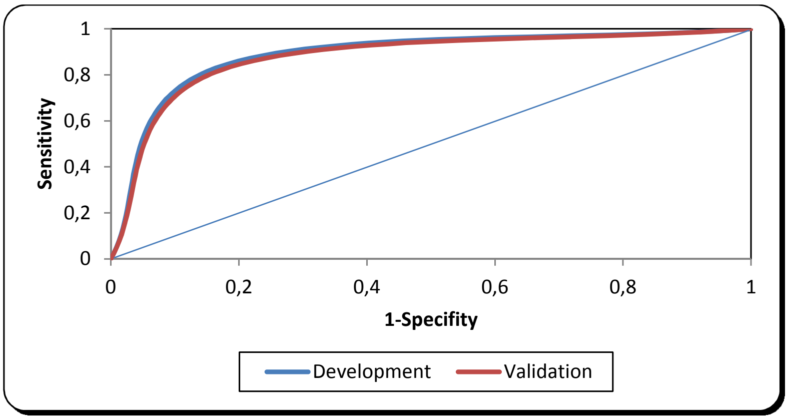
*Table B.2.6 Classification table results of propensity-to-buy CP2 model.*

| Classification table | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Prob | Correct | | Incorrect | | Percentage | | | | |
| Level | Event | Nonevent | Event | Nonevent | Correct | Sensitivity | Specificity | False POS | False NEG |
| 0.500 | 6742 | 835874 | 1339 | 168809 | 83.3 | 83.1 | 83.5 | 16.6 | 16.8 |

Developed model has represented correctness of 83.3% of observations on the 0.5 probability level. When target variable is 1 there are 6 742 observations which are corrected assigned and for target variable equals 0 there are 835 874 correct assignments. Moreover, for 16.6% of observations model gave the 0 value while it should be 1 value and for almost the same percentage, 16.8% it gave the 1 value while it should be 0 value.

Figure B.2.1 shows ROC curves presented for Development and Validation samples, based on the Sensitivity and Specificity values.

*Figure B.2.1 ROC charts for development and validation data sets of propensity-to-buy CP2 model.*




Based on the ROC curves it is easy to formulate the statement that created model fits the observations and it is going to predict marketing offer responders correctly.

Table B.2.7 shows the number of Events and Nonevents and Lift rate value for each decile of development sample and next table shows the same indicators but after applying the propensity-to-buy for CP2 model on the validation data set.

*Table B.2.7 Events and Nonevents results and Lift rate for application propensity-to-buy CP2 model on development sample.*

| Development at ALL | | | | | | |
|---|---|---|---|---|---|---|
| Decile | Non Resp. | Resp. | Total | Resp. Rate | Cum. Resp rate | % Lift |
| ~10% | 192 195 | 5 849 | 198 045 | 2,95% | 2,95% | 729,34% |
| ~20% | 199 137 | 1 109 | 200 245 | 0,55% | 1,75% | 431,39% |
| ~30% | 200 403 | 471 | 200 874 | 0,23% | 1,24% | 306,16% |
| ~40% | 199 972 | 273 | 200 245 | 0,14% | 0,96% | 237,89% |
| ~50% | 199 838 | 150 | 198 988 | 0,08% | 0,79% | 194,18% |
| ~60% | 202 646 | 114 | 202 760 | 0,06% | 0,66% | 163,74% |
| ~70% | 200 168 | 77 | 200 245 | 0,04% | 0,57% | 141,70% |
| ~80% | 200 206 | 39 | 200 245 | 0,02% | 0,50% | 124,59% |
| ~90% | 194 883 | 18 | 194 901 | 0,01% | 0,45% | 111,31% |
| ~100% | 205 893 | 11 | 205 904 | 0,01% | 0,40% | 100,00% |
| | 1 994 342 | 8 109 | 2 002 451 | 0,40% | | |

*Table B.2.8 Events and Nonevents results and Lift rate for application propensity-to-buy CP2 model on validation sample.*

| Validation at ALL | | | | | | |
|---|---|---|---|---|---|---|
| Decile | Non Resp. | Resp. | Total | Resp. Rate | Cum. Resp rate | % Lift |
| ~10% | 98 071 | 2 801 | 100 872 | 2,78% | 2,78% | 685,71% |
| ~20% | 98 097 | 507 | 98 604 | 0,51% | 1,66% | 409,52% |
| ~30% | 101 401 | 193 | 101 594 | 0,19% | 1,16% | 287,16% |
| ~40% | 99 718 | 163 | 99 881 | 0,16% | 0,91% | 225,66% |
| ~50% | 99 909 | 107 | 100 015 | 0,11% | 0,75% | 185,86% |
| ~60% | 97 848 | 94 | 97 942 | 0,10% | 0,65% | 159,32% |
| ~70% | 102 267 | 77 | 102 344 | 0,07% | 0,56% | 138,76% |
| ~80% | 98 863 | 39 | 98 902 | 0,04% | 0,50% | 122,81% |
| ~90% | 102 144 | 28 | 102 172 | 0,03% | 0,44% | 109,67% |
| ~100% | 98 854 | 47 | 98 901 | 0,05% | 0,40% | 100,00% |
| | 997 171 | 4 055 | 1 001 225 | 0,40% | | |

In the top decile of development sample there were 198 045 customers, with 5849 responders for a response rate of 2.95%. Compared to the average response rate of 0.40%, this gives a lift 7.29 (729%) for decile 1. Each successive decile has a lower response rate, what is correct since it means that model is ordering customers in a proper way. In the validation sample the results are fairly similar. In the top decile of validation sample there were 100872 customers, with 2801 responders and 2.78% of response rate. As average response rate is 0.40% Lift rate equals 6.86 (686%).

Figure B.2.2 presents the comparison of Lift rates for development and validation data sets.

*Figure B.2.2 Lift ratio results charts for development and validation data sets for propensity-to-buy for CP2 model.*

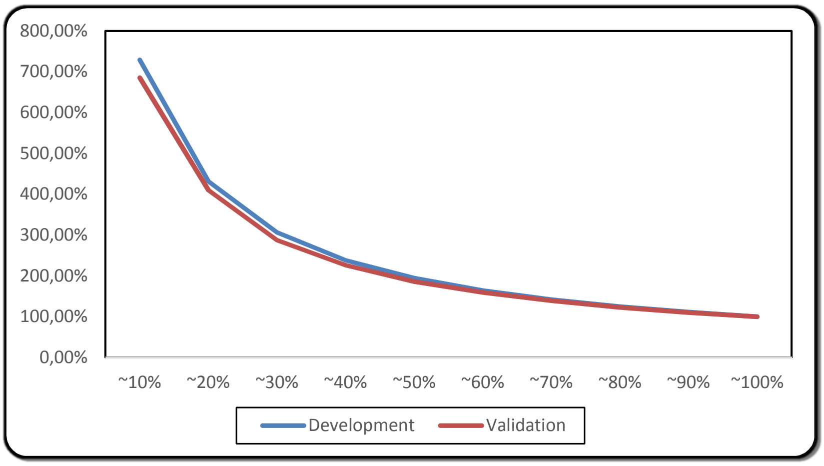



Chart compares the cumulative percent of responses captured as each decile is added to the target. According to the Table B.2.9 for development data set the top two deciles capture 85.80% of the responders (Events). This is compared to a random baseline where four deciles (40% of the population) would capture 40% of the responders. To remind - the greater the area between two lines – baseline and line with the information of cumulative percent of Responses Captured, the more the model is able to concentrate responders in the top deciles.

*Table B.2.9 Event and Nonevent distribution in deciles for development sample of propensity-to-buy for CP2 model.*

| Model Development at ALL | | | | | | | |
|---|---|---|---|---|---|---|---|
| Approx. score % | NonResp (NR) | Responders (R) | Prob. NonResp | % of all NonResp | % of all Resp | Cum. % NonResp | Cum. % Resp |
| ~10% | 192 195 | 5 849 | 0,9705 | 9,64% | 72,13% | 9,64% | 72,13% |
| ~20% | 199 137 | 1 109 | 0,9945 | 9,99% | 13,67% | 19,62% | 85,80% |
| ~30% | 200 403 | 471 | 0,9977 | 10,05% | 5,80% | 29,67% | 91,60% |
| ~40% | 199 972 | 273 | 0,9986 | 10,03% | 3,36% | 39,70% | 94,97% |
| ~50% | 198 838 | 150 | 0,9992 | 9,97% | 1,85% | 49,67% | 96,81% |
| ~60% | 202 646 | 114 | 0,9994 | 10,16% | 1,41% | 59,83% | 98,22% |
| ~70% | 200 168 | 77 | 0,9996 | 10,04% | 0,95% | 69,87% | 99,16% |
| ~80% | 200 206 | 39 | 0,9998 | 10,04% | 0,48% | 79,90% | 99,65% |
| ~90% | 194 883 | 18 | 0,9999 | 9,77% | 0,22% | 89,68% | 99,87% |
| ~100% | 205 893 | 11 | 0,9999 | 10,32% | 0,13% | 100,00% | 100,00% |
| Totals | 1 994 342 | 8 109 | | | | | |

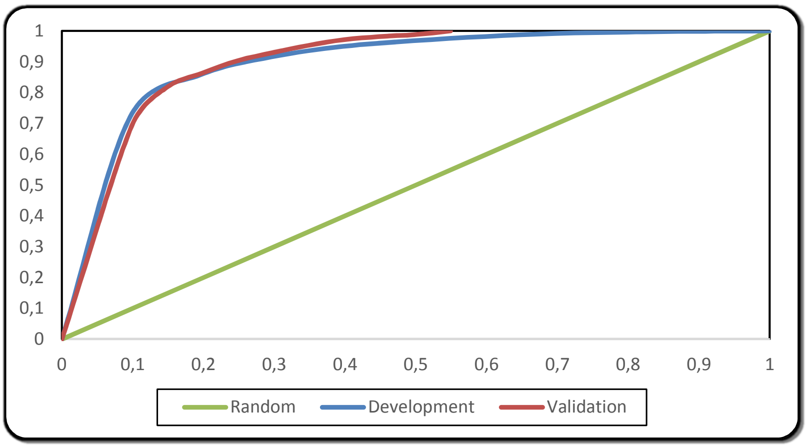
For validation data set the top two deciles capture 81.6% of responders what means that in this case the rule '80/20' is met and by selecting the customers who are qualified to first and second deciles it is possible to obtain 82% of response rate with save money of unsuccessful contact with 20% of prepared data base.

*Table B.2.10 Event and Nonevent distribution in deciles for validation sample of propensity-to-buy for CP2 model.*

| Model Validation at ALL | | | | | | | |
|---|---|---|---|---|---|---|---|
| Approx. score % | NonResp (NR) | Responders (R) | Prob. NonResp | % of all NonResp | % of all Resp | Cum. % NonResp | Cum. % Resp |
| ~10% | 98 071 | 2 801 | 0,9722 | 9,83% | 69,08% | 9,83% | 69,08% |
| ~20% | 98 097 | 507 | 0,9949 | 4,92% | 12,50% | 14,75% | 81,59% |
| ~30% | 101 401 | 193 | 0,9981 | 5,08% | 4,76% | 19,84% | 86,35% |
| ~40% | 99 718 | 163 | 0,9984 | 5,00% | 4,02% | 24,84% | 90,37% |
| ~50% | 99 909 | 107 | 0,9989 | 5,01% | 2,63% | 29,85% | 93,00% |
| ~60% | 97 848 | 94 | 0,999 | 4,91% | 2,31% | 34,75% | 95,30% |
| ~70% | 102 267 | 77 | 0,9993 | 5,13% | 1,89% | 39,88% | 97,19% |
| ~80% | 98 863 | 39 | 0,9996 | 4,96% | 0,96% | 44,84% | 98,15% |
| ~90% | 102 144 | 28 | 0,9997 | 5,12% | 0,69% | 49,96% | 98,84% |
| ~100% | 98 854 | 47 | 0,9995 | 4,96% | 1,16% | 54,92% | 100,00% |
| Totals | 997 171 | 4 055 | | | | | |

The Figure B.2.3 illustrates the numbers which are presented in the tables above.

*Figure B.2.3 Comparison of cumulative percent of Events and Nonevents for development and validation data set of propensity-to-buy for CP2 model.*




As a result it is worth to say that model represents good quality of fitness to the data.

## B.3 Propensity-to-buy Credit Product no 3 model - details

The goal of the third built model is to predict how much given customer is interested in buying CP3 within the marketing action.

CP3 is different in its characteristics from previous both products, as it depends on the portfolio which customer possesses. For this reason there was a need to build in the fact two propensity-to-buy models instead of only one. This treatment was regarded necessary since CP3 can be sold with or without Deposit Product no 3, and this fact should be taken into account while building the model. Moreover, the customer who bought CP3 separately (but first he/she had to buy Deposit Product no 3) is considered to be different from one who bought it in a package with Deposit Product no 3 in the same, or almost the same time. If difference between dates of purchase of 2 products is up to 30 days, it is treated like purchase the products in the same time. There are some reasons of such situation: there are some delays in the bank systems and processes or the customer can forget to ask for another product or sometimes it is impossible to buy or register 2 products in the same date, even if the intention was different.

As in the previous cases, but this time connected with CP3, there were some marketing campaigns which took place whole year before the time of building models. However, they were addressed only these customers who had already bought Deposit Product no 3. In opposite case, if company wants to launch the campaign dedicated to customers without Deposit Product no 3 it is necessary to offer them 2 products: CP3 and also Deposit Product no 3. That is the reason why this kind of model focuses on customers who bought Credit Product no 3 with Deposit Product no 3 without any CRM stimulation, only because of their own needs (in assumption). For the valid understanding the first case, when customer has purchased the Deposit Product no 3 before the Credit Product no 3 model will be called ‘propensity-to-buy for CP3A model’. The second option will be called ‘propensity-to-buy for CP3B model’.

As far as for CP3A propensity-to-buy model is concerned, there were six main campaigns with mailing contact and four followers campaigns with phone contact in the input data set. However, there was such a little number of successes available (customers who were interested in purchasing this product), that it was decided that it is necessary to create the model based on all purchases rather than only on the marketing campaigns` successes. Observation window, taking into account 14 months before building the model, in the numbers included 9 743 customers who bought Credit Product no 3, having bought Deposit Product no 3 minimum 30 days before that. Since created data set contains also customers who were not offered in the marketing actions a data set of 36 890 customers, who do not have Credit Product no 3 but they do have Deposit Product no 3, were randomly selected by sampling without replacement put of the whole set 424 224 customers. After union of these two data sets first part is Event data set and second one is Nonevent data set. It gives response rate equals 20.89%, which normally is not possible to meet and it equals 2.30%. Response Profile is shown like below:

*Table B.3.1 Response Profile for CP3A.*

Response Profile

| Ordered Value | CP1 | Total Frequency |
|---|---|---|
| 1 | Event \| 1 | 9 743 |
| 2 | NonEvent \| 0 | 424 224 |

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied

If propensity-to-buy for CP3B model is taken into account there were 1431 customers who bought Credit Product no 3 with Deposit Product no 3 within the 14 months. This part represents target variable=1. Nonevents part in number of 7 240 was randomly selected by sampling without replacement from customers who did not buy either CP3 or Deposit Product no 3, out of 853 990 customers. This data set included response rate at level of 16.15%

what is also normally very difficult to meet, even for other products, it equals 0,17%. Below there is Response Profile Table placed:

*Table B.3.2 Response Profile for CP3B.*

Response Profile

| Ordered Value | CP1 | Total Frequency |
|---|---|---|
| 1 | Event \| 1 | 1 431 |
| 2 | NonEvent \| 0 | 853 990 |

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied

In case of this propensity-to-buy model, number of Events is very small. The number of Nonevents is small as well. Random sampling with no replacement which is a way to create Nonevent data set is penetrated only one-hundredth of the whole available data base of customers. In order to obtain reliable results which can be applied on this whole data base the solution of ten development samples constructing was introduced. As it was done in previous models, after created ten propensity-to-buy models the best propensity to buy model was chosen.

Willingness of purchasing Credit Product no 3 can be described with following formula:

$$\begin{cases} P_{CP3A} = Prob(Y_{CP3A}|X_{CP3A}) = \dfrac{1}{1+e^{-L_{CP3A}}} & ,if\ CP3_{vDP3} > 0 \\ P_{CP3B} = Prob(Y_{CP3B}|X_{CP3A}) = \dfrac{1}{1+e^{-L_{CP3B}}} & ,if\ CP3_{vDP3} = 0 \end{cases} \tag{B.3.1}$$

where

$$\begin{aligned} L_{CP3A} &= \alpha + \beta_1 X_{1CP3A} + \beta_2 X_{2CP3A} + \cdots + \beta_k X_{kCP3A} \\ L_{CP3B} &= \alpha + \beta_1 X_{1CP3B} + \beta_2 X_{2CP3B} + \cdots + \beta_k X_{kCP3B} \end{aligned} \tag{B.3.2}$$

where *CP3A* means Credit Product no 3A, *CP3B* means Credit Product no 3B, *CP3* means Credit Product no 3, $CP3_{vDP3}$ means Deposit Product no 3. The Tables below (B.3.1, B.3.2) list the parameter estimates, their standard errors and the results of the Wald test for individual parameters for models created for willingness to buy Credit Product no 3.

*Table B.3.1 Analysis of Maximum Likelihood Estimates for CP3A.*

| Analysis of Maximum Likelihood Estimates [CP3A] | | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Parameter type | Parameter description | DF | Estimator | Standard Error | Wald Chi-Square | Pr > Chi-Sq. |
| CP3A_v1 | account-related | number of active credit products no 6 which customer has got in other banks | 1 | -0.2727 | 0.0309 | 77.9526 | <.0001 |
| CP3A_v2 | account-related | number of active credit products no 1 which customer has got in other banks | 1 | -0.3046 | 0.0249 | 150.1408 | <.0001 |
| CP3A_v3 | account-related | number of active credit products no 2 which customer has got in other banks | 1 | -0.2983 | 0.0293 | 103.9423 | <.0001 |
| CP3A_v4 | account-related | number of days of customer delay in payments | 1 | -0.00020 | 0.000017 | 146.8322 | <.0001 |
| CP3A_v5 | CRM-related | customer was communicated via phone in last 3 months | 1 | 0.6086 | 0.0481 | 159.8392 | <.0001 |
| CP3A_v6 | CRM-related | customer was communicated via phone in last 6 months | 1 | 0.1828 | 0.0434 | 17.7520 | <.0001 |
| CP3A_v7 | CRM-related | customer has bought credit product no 1 in last 3 months in CRM campaign | 1 | 0.8660 | 0.3316 | 6.8216 | 0.0090 |
| CP3A_v8 | CRM-related | customer has bought any CRM product in last 3 months | 1 | -0.7858 | 0.2178 | 13.0236 | 0.0003 |
| CP3A_v9 | CRM-related | customer has bought any CRM product in last 6 months | 1 | 0.5039 | 0.1208 | 17.4091 | <.0001 |
| CP3A_v10 | demographic dummy | number of months of being bank customer | 1 | -0.00521 | 0.000325 | 257.1029 | <.0001 |
| CP3A_v11_d1 | demographic dummy | customer does not have any information about employment | 1 | -2.8470 | 0.0706 | 1625.1314 | <.0001 |
| CP3A_v11_d10 | demographic dummy | customer job is like 'other' | 1 | -2.5445 | 0.1047 | 590.2927 | <.0001 |
| CP3A_v11_d2 | demographic dummy | customer works on full time | 1 | 0.7694 | 0.0534 | 207.3724 | <.0001 |
| CP3A_v11_d5 | demographic dummy | customer is retired | 1 | 1.3288 | 0.0667 | 396.9053 | <.0001 |
| CP3A_v12_d3 | demographic dummy | customer marital status is married | 1 | -0.3788 | 0.0368 | 105.9753 | <.0001 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CP3A_v12_d6 | demographic dummy | customer does not have information about marital status | 1 | 0.4729 | 0.1288 | 13.4910 | 0.0002 |
| CP3A_v13_d1 | demographic dummy | customer does not have any information about the income | 1 | -0.8341 | 0.0511 | 266.6849 | <.0001 |
| CP3A_v13_d2 | demographic dummy | customer income is >=10 and <1236) | 1 | -1.3407 | 0.0594 | 509.7710 | <.0001 |
| CP3A_v13_d3 | demographic dummy | customer income is >=1236 and <1775 | 1 | -0.3660 | 0.0457 | 64.0570 | <.0001 |
| CP3A_v14_d1 | demographic dummy | customer does not have any information about the occupation class | 1 | 0.9063 | 0.0514 | 310.3762 | <.0001 |
| CP3A_v14_d2 | demographic dummy | customer occupation class is white-collar worker | 1 | 0.1185 | 0.0468 | 6.4143 | 0.0113 |
| CP3A_v14_d5 | demographic dummy | customer occupation class is director | 1 | 0.7587 | 0.0757 | 100.3936 | <.0001 |
| CP3A_v15 | product characteristics | customer has got credit product no 1 | 1 | 0.2773 | 0.0350 | 62.7370 | <.0001 |
| CP3A_v16 | product characteristics | customer has got deposit product no 2 | 1 | 0.8068 | 0.0408 | 390.6447 | <.0001 |
| CP3A_v17 | product characteristics | sum of the credit products no 2 | 1 | 0.1189 | 0.0144 | 67.7686 | <.0001 |
| CP3A_v18 | product characteristics | customer has got deposit product no 1 | 1 | 0.3593 | 0.0389 | 85.2666 | <.0001 |
| CP3A_v19_d34 | product characteristics dummy | customer has got (2;4 > different products | 1 | 0.1033 | 0.0399 | 6.7121 | 0.0096 |
| CP3A_v19_d78 | product characteristics dummy | customer has got (8, 26> different products | 1 | -0.1619 | 0.0390 | 17.2800 | <.0001 |
| CP3A_v0 | | | 1 | -0.3647 | 0.0667 | 29.8685 | <.0001 |

All the estimated parameters are significantly different from zero ($p<0.05$). It simply means that all used variables could be interpreted and they explain the estimated propensity-to-buy CP 3A not only in the marketing campaign, but in general. Variables ending with 'd' means that this variable is dummy variable of the original one. Variables CP3A_v10-v14 are customer-related and they describe customer from demographic point of view mostly, variables CP3A_v5-v9 are linked with CRM campaigns history and frequency and way of contact with customer. Variables CP3A_v15-v19 say about the history of the customer portfolio and CP3A_v1-v4 give the picture of customer credit background in other banks and his delays with the credit repayments.

*Table B.3.2 Analysis of Maximum Likelihood Estimates for CP3B.*

| Analysis of Maximum Likelihood Estimates [CP3B] | | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Parameter type | Parameter description | DF | Estimator | Standard Error | Wald Chi-Square | Pr > Chi-Sq. |
| CP3B_v1 | CRM-related | customer was communicated via phone in last 6 months | 1 | 0.3165 | 0.0932 | 11.5178 | 0.0007 |
| CP3B_v2 | CRM-related | customer has bought credit product no 1 in bank campaign in last year | 1 | 1.5384 | 0.5799 | 7.0383 | 0.0080 |
| CP3B_v3 | CRM-related | customer has bought credit product no 1 in bank campaign | 1 | 0.3784 | 0.1595 | 5.6242 | 0.0177 |
| CP3B_v4 | demographic | kind of customer employer agreement type | 1 | 0.1959 | 0.0205 | 91.4869 | <.0001 |
| CP3B_v5 | demographic | kind of customer marital status | 1 | 0.4266 | 0.1313 | 10.5521 | 0.0012 |
| CP3B_v5_d1 | demographic dummy | no information about marital status | 1 | 2.8854 | 0.4547 | 40.2748 | <.0001 |
| CP3B_v5_d4 | demographic dummy | customer is divorced or customer is a widow/widower | 1 | -0.9436 | 0.3766 | 6.2796 | 0.0122 |
| CP3B_v6 | demographic | kind of customer occupation class | 1 | 0.3640 | 0.0697 | 27.2778 | <.0001 |
| CP3B_v6_d4 | demographic dummy | customer occupation class is manager/owner | 1 | -0.7284 | 0.2333 | 9.7501 | 0.0018 |
| CP3B_v6_d5 | demographic dummy | customer occupation class is other | 1 | -2.0308 | 0.2844 | 51.0035 | <.0001 |
| CP3B_v6_d6 | demographic dummy | customer occupation class is not mentioned | 1 | -1.4774 | 0.4157 | 12.6314 | 0.0004 |
| CP3B_v7 | demographic | kind of customer education | 1 | 0.1955 | 0.0392 | 24.8869 | <.0001 |
| CP3B_v8_d1 | demographic dummy | no information about employer type | 1 | -0.2414 | 0.0410 | 34.7029 | <.0001 |
| CP3B_v8_d10 | demographic dummy | customer works on part time | 1 | -3.8614 | 0.3127 | 152.4805 | <.0001 |
| CP3B_v8_d2 | demographic dummy | customer works on not full time | 1 | 1.6474 | 0.1444 | 130.1517 | <.0001 |
| CP3B_v8_d5 | demographic dummy | customer is retiring | 1 | 1.6938 | 0.1877 | 81.4363 | <.0001 |
| CP3B_v9_d2 | demographic dummy | customer income >0 and <807 PLN | 1 | -3.1816 | 0.1935 | 270.3832 | <.0001 |
| CP3B_v9_d3 | demographic dummy | customer income >807 and <1100 PLN | 1 | -2.4165 | 0.2295 | 110.8192 | <.0001 |
| CP3B_v9_d4 | demographic dummy | customer income >1100 and <1848 PLN | 1 | -0.8383 | 0.1066 | 61.8422 | <.0001 |
| CP3B_v9_d7 | demographic dummy | customer income >3134 and <73363 PLN | 1 | 0.4559 | 0.1208 | 14.2552 | 0.0002 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CP3B_v10_d1 | demographic dummy | customer does not have dependent on income people | 1 | -0.8974 | 0.2888 | 9.6584 | 0.0019 |
| CP3B_v10_d2 | demographic dummy | customer has one dependent on income person | 1 | 0.7171 | 0.1586 | 20.4400 | <.0001 |
| CP3B_v11 | products characteristics | customer has got credit product no 1 | 1 | 0.6171 | 0.1155 | 28.5610 | <.0001 |
| CP3B_v12 | products characteristics | customer has got group of deposit products no 1 | 1 | 1.1230 | 0.1559 | 51.8766 | <.0001 |
| CP3B_v13 | products characteristics | customer has got deposit product no 2 | 1 | 1.5762 | 0.3164 | 24.8248 | <.0001 |
| CP3B_v14 | products characteristics | sum of the credit product no 2 | 1 | 1.3355 | 0.0983 | 184.6730 | <.0001 |
| CP3B_v15 | products characteristics | sum of credit group of products no 1 | 1 | -0.1613 | 0.0276 | 34.2754 | <.0001 |
| CP3B_v16 | products characteristics | sum of credit group of products no 2 | 1 | -0.8956 | 0.0755 | 140.6766 | <.0001 |
| CP3B_v17_d2 | products characteristics dummy | customer has got <0;1) different products | 1 | -0.6119 | 0.1317 | 21.5783 | <.0001 |
| CP3B_v17_d5 | products characteristics dummy | customer has got <3;5) different products | 1 | 0.2972 | 0.1221 | 5.9251 | 0.0149 |
| CP3B_v17_d6 | products characteristics dummy | customer has got <5;10) different products | 1 | 0.8121 | 0.1327 | 37.4589 | <.0001 |
| CP3B_v17_d7 | products characteristics dummy | customer has got >=10 different products | 1 | 1.3109 | 0.2294 | 32.6647 | <.0001 |
| CP3B_v0 | | | 1 | -3.8948 | 0.3390 | 132.0234 | <.0001 |

All the estimated parameters are significantly different from zero ($p<0.05$). It simply means that all used variables could be interpreted and they explain the estimated propensity-to-buy CP3B in general. Variables ending with 'd' means that this variable is dummy variable of the original one. Variables CP3B_v4-v10 are customer-related and they describe customer from demographic point of view mainly, variables CP3B_v1-v3 are linked with the CRM campaigns history and frequency and the way of contact with the customer. Variables CP2_v11-v17 say about the history of the customer.

The Association of Predicted Probabilities and Observed Responses results, which are showed by Table B.3.3 display measures of association between predicted probabilities and observed responses which include a breakdown of the number of pairs with different responses and four rank correlation indexes. After launched the model on development sample of CP3A there were 89.4% of Concordant pairs, meaning for 89% of 3,59E+08 pairs model estimated correctly higher probability for higher ordered value than for lower ordered value. For 10.3% of total number of pairs model gave higher probability for lower ordered value and for only 0.3% pairs probability for both ordered values was the same.

*Table B.3.3 Association of Predicted Probabilities and Observed Responses for CP3A.*

| Association of Predicted Probabilities and Observed Responses [CP3A] | | | |
|---|---|---|---|
| Percent Concordant | 89.4 | Somers' D | 0.791 |
| Percent Discordant | 10.3 | Gamma | 0.793 |
| Percent Tied | 0.3 | Tau-a | 0.261 |
| Pairs | 359409527 | c | 0.895 |

For propensity-to-buy for CP3B model the results are displayed by Table B.3.4. In this case there were 94% of Concordant pairs, meaning for up to 94% of 1,06E+07 pairs model estimated correctly higher probability for higher ordered value than for lower ordered value. For 5.8% of total number of pairs model gave higher probability for lower ordered value and for only 0.2% pairs probability for both ordered values was the same.

*Table B.3.4 Association of Predicted Probabilities and Observed Responses for CP3B.*

| Association of Predicted Probabilities and Observed Responses [CP3B] | | | |
|---|---|---|---|
| Percent Concordant | 94.0 | Somers' D | 0.882 |
| Percent Discordant | 5.8 | Gamma | 0.883 |
| Percent Tied | 0.2 | Tau-a | 0.239 |
| Pairs | 10626606 | c | 0.941 |

Odds ratios estimates for parameters used in propensity-to-buy for CP3A model are able to find below.

*Table B.3.3.5 Odds ratio estimates for CP3A.*

| Odds Ratio Estimates [CP3A] | | | | |
|---|---|---|---|---|
| Effect | Parameter description | Point Estimate | 95% Wald Confidence Limits | |
| CP3A_v1 | number of active credit products no 6 which customer has got in other banks | 2.158 | 1.944 | 2.397 |
| CP3A_v2 | number of active credit products no 1 which customer has got in other banks | 0.058 | 0.051 | 0.067 |
| CP3A_v3 | number of active credit products no 2 which customer has got in other banks | 0.434 | 0.393 | 0.480 |
| CP3A_v4 | number of days of customer delay in payments | 2.475 | 2.238 | 2.738 |
| CP3A_v5 | customer was communicated via phone in last 3 months | 3.777 | 3.314 | 4.304 |
| CP3A_v6 | customer was communicated via phone in last 6 months | 1.838 | 1.672 | 2.020 |
| CP3A_v7 | customer has bought credit product no 1 in last 3 months in CRM campaign | 1.201 | 1.103 | 1.307 |
| CP3A_v8 | customer has bought any CRM product in last 3 months | 2.241 | 2.068 | 2.427 |
| CP3A_v9 | customer has bought any CRM product in last 6 months | 2.136 | 1.841 | 2.477 |
| CP3A_v10 | number of months of being bank customer | 1.126 | 1.095 | 1.159 |
| CP3A_v11_d1 | customer does not have any information about employment | 1.126 | 1.027 | 1.234 |
| CP3A_v11_d10 | customer job is like 'other' | 0.685 | 0.637 | 0.736 |
| CP3A_v11_d2 | customer works on full time | 0.079 | 0.064 | 0.096 |
| CP3A_v11_d5 | customer is retired | 1.655 | 1.306 | 2.097 |
| CP3A_v12_d3 | customer marital status is married | 1.605 | 1.247 | 2.065 |
| CP3A_v12_d6 | customer does not have information about marital status | 0.694 | 0.634 | 0.759 |
| CP3A_v13_d1 | customer does not have any information about the income | 1.320 | 1.232 | 1.413 |
| CP3A_v13_d2 | customer income is >=10 and <1236) | 0.850 | 0.788 | 0.918 |
| CP3A_v13_d3 | customer income is >=1236 and <1775 | 1.432 | 1.327 | 1.546 |
| CP3A_v14_d1 | customer does not have any information about the occupation class | 0.262 | 0.233 | 0.294 |
| CP3A_v14_d2 | customer occupation class is white-collar worker | 0.456 | 0.297 | 0.698 |
| CP3A_v14_d5 | customer occupation class is director | 1.109 | 1.025 | 1.199 |
| CP3A_v15 | customer has got credit product no 1 | 1.000 | 1.000 | 1.000 |
| CP3A_v16 | customer has got deposit product no 2 | 0.995 | 0.994 | 0.995 |
| CP3A_v17 | sum of the credit products no 2 | 2.377 | 1.241 | 4.553 |
| CP3A_v18 | customer has got deposit product no 1 | 0.737 | 0.702 | 0.774 |
| CP3A_v19_d34 | customer has got (2;4 > different products | 0.742 | 0.701 | 0.786 |
| CP3A_v19_d78 | customer has got (8, 26> different products | 0.761 | 0.717 | 0.809 |

For variable CP3A_v6, where odds ratio=1.838 it can be interpreted as follows: if there are customers who are different from each other only in terms of being communicated via phone in last 6 months, the customer, who was communicated via phone in last 6 months has an odds ratio of buying CP3A versus not buying CP3A which is 83,8 higher than the customer who was not communicated. For variable CP3A_v18, where odds ratio=0.737 it can be described in a way: if there are customers who are different from each other only in terms of having deposit product no 1, the customer who has got this product has ann odds ratio of buying CP3A versus not buying which is 26,3% lower than the customer who does not have this Deposit Product no 1.

*Table B.3.6 Odds ratio estimates for CP3B.*

| Odds Ratio Estimates [CP3B] | | | | |
|---|---|---|---|---|
| Effect | Parameter description | Point Estimate | 95% Wald Confidence Limits | |
| CP3B_v1 | customer was communicated via phone in last 6 months | 1.372 | 1.143 | 1.647 |
| CP3B_v2 | customer has bought credit product no 1 in bank campaign in last year | 4.657 | 1.495 | 14.510 |
| CP3B_v3 | customer has bought credit product no 1 in bank campaign | 1.460 | 1.068 | 1.996 |
| CP3B_v4 | kind of customer employer agreement type | 1.216 | 1.169 | 1.266 |
| CP3B_v5 | kind of customer marital status | 1.532 | 1.184 | 1.982 |
| CP3B_v5_d1 | no information about marital status | 17.910 | 7.347 | 43.662 |
| CP3B_v5_d4 | customer is divorced or customer is a widow/widower | 0.389 | 0.186 | 0.814 |
| CP3B_v6 | kind of customer occupation class | 1.439 | 1.255 | 1.650 |
| CP3B_v6_d4 | customer occupation class is manager/owner | 0.483 | 0.306 | 0.762 |
| CP3B_v6_d5 | customer occupation class is other | 0.131 | 0.075 | 0.229 |
| CP3B_v6_d6 | customer occupation class is not mentioned | 0.228 | 0.101 | 0.515 |
| CP3B_v7 | kind of customer education | 1.216 | 1.126 | 1.313 |
| CP3B_v8_d1 | no information about employer type | 0.786 | 0.725 | 0.851 |
| CP3B_v8_d10 | customer works on part time | 0.021 | 0.011 | 0.039 |
| CP3B_v8_d2 | customer works on not full time | 5.193 | 3.913 | 6.892 |
| CP3B_v8_d5 | customer is retired | 5.440 | 3.766 | 7.859 |
| CP3B_v9_d2 | customer income >0 and <807 PLN | 0.042 | 0.028 | 0.061 |
| CP3B_v9_d3 | customer income >807 and <1100 PLN | 0.089 | 0.057 | 0.140 |
| CP3B_v9_d4 | customer income >1100 and <1848 PLN | 0.432 | 0.351 | 0.533 |
| CP3B_v9_d7 | customer income >3134 and <73363 PLN | 1.578 | 1.245 | 1.999 |
| CP3B_v10_d1 | customer does not have dependent on income people | 0.408 | 0.231 | 0.718 |
| CP3B_v10_d2 | customer has one dependent on income person | 2.049 | 1.501 | 2.796 |
| CP3B_v11 | customer has got credit product no 1 | 1.853 | 1.478 | 2.324 |
| CP3B_v12 | customer has got group of deposit products no 1 | 3.074 | 2.265 | 4.173 |
| CP3B_v13 | customer has got deposit product no 2 | 4.837 | 2.602 | 8.991 |
| CP3B_v14 | sum of the credit product no 2 | 3.802 | 3.136 | 4.609 |
| CP3B_v15 | sum of credit group of products no 1 | 0.851 | 0.806 | 0.898 |
| CP3B_v16 | sum of credit group of products no 2 | 0.408 | 0.352 | 0.473 |
| CP3B_v17_d2 | customer has got <0;1) different products | 0.542 | 0.419 | 0.702 |

| | | | | |
|---|---|---|---|---|
| CP3B_v17_d5 | customer has got <3;5) different products | 1.346 | 1.060 | 1.710 |
| CP3B_v17_d6 | customer has got <5;10) different products | 2.253 | 1.737 | 2.922 |
| CP3B_v17_d7 | customer has got >=10 different products | 3.710 | 2.366 | 5.815 |

In the propensity-to-buy CP3B model there is no variable for which odds ratio equal almost 1. It means that such variable has got no association with the target variable. However there are variables with odds ratio value greater than 1 (19 variables) or smaller than 1 (13 variables). For variable CP3B_v17_d5, where odds ratio=1.346 it can be interpreted as follows: if there are customers who are different from each other only in terms of having different products in portfolio, the customer, who has got <3;5) different products in his/her portfolio has an odds ratio of buying CP3B versus not buying which is 34,6% lower than the customer who has got less or more different products in portfolio. For variable CP3B_v5_d4, where odds ratio=0.389 it can be described in a way: if there are customers who are different from each other only in terms of marital status, the customer who is divorced or who is widow or widower has about 61,1% less chances to have high probability to buy CP3B than the customer who are not divorced and are not widow/widower.

The Table B.3.7 and the Table B.3.8 illustrate the Hosmer and Lemeshow test results of propensity-to-buy Credit Product no 3 model, which examined model fit to the observed values. In these two models there is no reason to reject the null hypothesis, what can allow to conclude that developed models are well suited to the data.

*Table B.3.7 The Hosmer-Lemeshow goodness-of-fit results of propensity-to-buy CP3A model.*

| Hosmer-Lemeshow test results | | | | | |
|---|---|---|---|---|---|
| Group | Total | Target_variable=1 | | Target_variable=0 | |
| | | Observed | Predicted | **Observed** | Predicted |
| 1 | 43 413 | 62 | 36 | 43 351 | 43 378 |
| 2 | 43 395 | 150 | 97 | 43 245 | 43 298 |
| 3 | 43 395 | 159 | 136 | 43 236 | 43 259 |
| 4 | 43 395 | 144 | 167 | 43 251 | 43 228 |
| 5 | 43 395 | 201 | 206 | 43 194 | 43 189 |
| 6 | 43 395 | 342 | 326 | 43 053 | 43 069 |
| 7 | 43 395 | 529 | 714 | 42 866 | 42 681 |
| 8 | 43 395 | 1 600 | 1 661 | 41 795 | 41 734 |
| 9 | 43 395 | 2 840 | 2 667 | 40 555 | 40 728 |
| 10 | 43 395 | 3 716 | 3 733 | 39 679 | 39 661 |

*Table B.3.8 The Hosmer-Lemeshow goodness-of-fit results of propensity-to-buy CP3B model.*

| Hosmer-Lemeshow test results | | | | | |
|---|---|---|---|---|---|
| Group | Total | Target_variable=1 | | Target_variable=0 | |
| | | Observed | Predicted | Observed | Predicted |
| 1 | 85 571 | 2 | 0 | 85 569 | 85 571 |
| 2 | 85 764 | 1 | 1 | 85 763 | 85 763 |
| 3 | 85 764 | 7 | 3 | 85 757 | 85 761 |
| 4 | 85 668 | 7 | 7 | 85 661 | 85 660 |
| 5 | 85 571 | 11 | 14 | 85 560 | 85 557 |
| 6 | 85 571 | 32 | 29 | 85 539 | 85 543 |
| 7 | 85 571 | 66 | 67 | 85 505 | 85 504 |
| 8 | 85 571 | 159 | 178 | 85 412 | 85 393 |
| 9 | 85 571 | 415 | 404 | 85 156 | 85 167 |
| 10 | 84 798 | 731 | 727 | 84 067 | 84 071 |

Next tables display the Classification results of the created propensity-to-buy for CP 3 models.

*Table B.3.9 Classification table results of propensity-to-buy CP3A model.*

| Classification table [CP3A] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Prob | Correct | | Incorrect | | Percentage | | | | |
| Level | Event | Nonevent | Event | Nonevent | Correct | Sensitivity | Specificity | False POS | False NEG |
| 0.500 | 6 154 | 396 198 | 2 437 | 140 070 | 87.1 | 63.2 | 93.4 | 28.4 | 9.4 |

Developed model has represented correctness of 87.1% of observations at the 0.5 probability level. When target variable is 1 there are 6154 observations which are corrected assigned and for target variable equals 0 there are 396 198 correct assignments. Moreover, for 28.4% of observations model gave the 0 value while it should be 1 value and for 9.4% it gave the 1 value while it should be 0 value.

*Table B.3.10 Classification table results of propensity-to-buy CP3B model.*

| Classification table [CP3B] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Prob | Correct | | Incorrect | | Percentage | | | | |
| Level | Event | Nonevent | Event | Nonevent | Correct | Sensitivity | Specificity | False POS | False NEG |
| 0.500 | 907 | 821 905 | 279 | 196 650 | 90.9 | 63.4 | 96.2 | 23.5 | 6.8 |

Developed model has represented correctness of 90.9% of observations at the 0.5 probability level. When target variable is 1 there are 907 observations which are corrected assigned and for target variable equals 0 there are 196 650 correct assignments. Moreover, for 23.5% of observations model gave the 0 value while it should be 1 value and for 6.8% it gave the 1 value while it should be 0 value.

Below figures present ROC curves presented. They show model fit for Development and Validation samples, based on Sensitivity and Specificity values for the both presented cases.

*Figure B.3.1 ROC charts for development and validation data sets of propensity-to-buy CP3A model.*

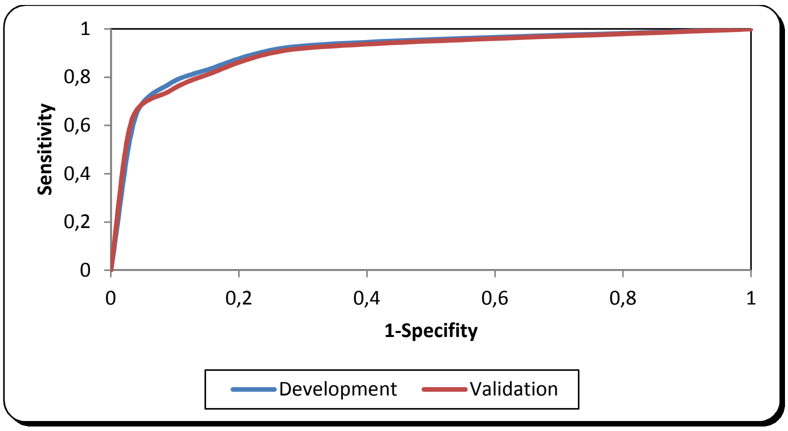
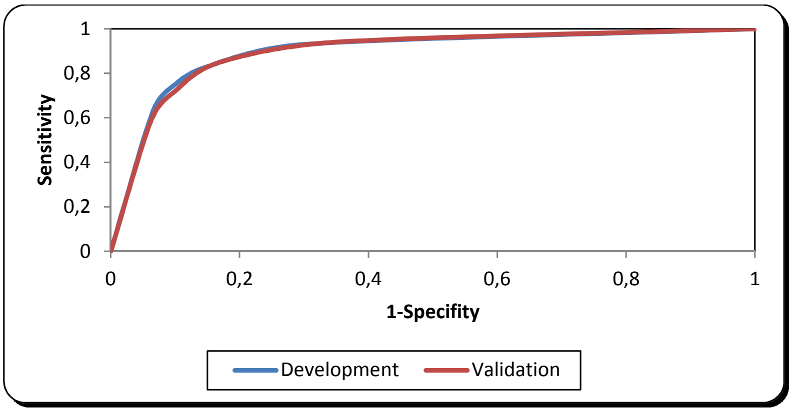


*Figure B.3.2 ROC charts for development and validation data sets of propensity-to-buy CP3B model.*




The ROC curves confirm the fact that created models fit the observations and they are going to predict marketing offer responders correctly.

Table B.3.11 and Table B.3.12 show the number of Events and Nonevents and Lift rate value for each decile of Development samples. Table B.3.13 and Table B.3.14 show the same indicators but after applying the propensity-to-buy for Credit Product no 3 models on the Validation data sets.

*Table B.3.11 Events and Nonevents results and Lift rate for application propensity-to-buy CP3A model on development sample.*

| Development at ALL [CP3A] | | | | | | |
|---|---|---|---|---|---|---|
| Decile | Non Resp. | Resp. | Total | Resp. Rate | Cum. Resp rate | % Lift |
| ~10% | 39 679 | 3 716 | 43 395 | 8,56% | 8,56% | 381,42% |
| ~20% | 40 555 | 2 840 | 43 395 | 6,54% | 7,55% | 336,46% |
| ~30% | 41 795 | 1 600 | 43 395 | 3,69% | 6,26% | 279,05% |
| ~40% | 42 866 | 529 | 43 395 | 1,22% | 5,00% | 222,86% |
| ~50% | 43 053 | 342 | 43 395 | 0,79% | 4,16% | 185,31% |
| ~60% | 43 194 | 201 | 43 395 | 0,46% | 3,54% | 157,86% |
| ~70% | 43 251 | 144 | 43 395 | 0,33% | 3,09% | 137,42% |
| ~80% | 43 236 | 159 | 43 395 | 0,37% | 2,75% | 122,29% |
| ~90% | 43 245 | 150 | 43 395 | 0,35% | 2,48% | 110,41% |
| ~100% | 43 351 | 62 | 43 413 | 0,14% | 2,25% | 100,00% |
| | 424 224 | 9 743 | 433 967 | 2,25% | | |

*Table B.3.12 Events and Nonevents results and Lift ratefor application propensity-to-buy CP3A model on validation sample.*

| Validation at ALL [CP3A] | | | | | | |
|---|---|---|---|---|---|---|
| Decile | Non Resp. | Resp. | Total | Resp. Rate | Cum. Resp rate | % Lift |
| ~10% | 32 094 | 712 | 32 806 | 2,17% | 2,17% | 445,74% |
| ~20% | 18 997 | 153 | 19 150 | 0,80% | 1,66% | 341,93% |
| ~30% | 23 472 | 125 | 23 597 | 0,53% | 1,31% | 269,12% |
| ~40% | 53 462 | 258 | 53 720 | 0,48% | 0,97% | 198,27% |
| ~50% | 43 493 | 130 | 43 623 | 0,30% | 0,80% | 163,69% |
| ~60% | 53 821 | 74 | 53 895 | 0,14% | 0,64% | 131,49% |
| ~70% | 46 611 | 82 | 46 693 | 0,18% | 0,56% | 115,20% |
| ~80% | 29 545 | 47 | 29 592 | 0,16% | 0,52% | 107,14% |
| ~90% | 11 464 | 10 | 11 474 | 0,09% | 0,51% | 103,88% |
| ~100% | 15 480 | 16 | 15 496 | 0,10% | 0,49% | 100,00% |
| | 328 438 | 1 607 | 330 045 | 0,49% | | |

In the top first decile of development sample there were 43 395 customers, where there were 3 716 responders for a response rate of 8.56%. Compared to the average response rate of 2.25%, this gives a lift 3.81 (381%) for decile 1. In the validation sample the results are quite similar, but Lift rate is better for the two first deciles. In the top decile of validation sample there were 3 280 663 customers, with 712 responders and 2,17% of response rate. As average response rate is 0,49%. Lift rate equals 4.46 (446%) already, what is better alike development sample.

*Table B.3.13 Events and Nonevents results and Lift rate for application propensity-to-buy CP3B model on development sample.*

| Development at ALL [CP3B] | | | | | | |
|---|---|---|---|---|---|---|
| Decile | Non Resp. | Resp. | Total | Resp. Rate | Cum. Resp rate | % Lift |
| ~10% | 84 067 | 731 | 84 798 | 0,86% | 0,86% | 515,31% |
| ~20% | 85 156 | 415 | 85 571 | 0,48% | 0,67% | 402,10% |
| ~30% | 85 412 | 159 | 85 571 | 0,19% | 0,51% | 304,80% |
| ~40% | 85 505 | 66 | 85 571 | 0,08% | 0,40% | 239,98% |
| ~50% | 85 539 | 32 | 85 571 | 0,04% | 0,33% | 196,37% |
| ~60% | 85 560 | 11 | 85 571 | 0,01% | 0,28% | 164,88% |
| ~70% | 85 661 | 7 | 85 668 | 0,01% | 0,24% | 141,97% |
| ~80% | 85 757 | 7 | 85 764 | 0,01% | 0,21% | 124,78% |
| ~90% | 85 763 | 1 | 85 764 | 0,00% | 0,19% | 110,96% |
| ~100% | 85 569 | 2 | 85 571 | 0,00% | 0,17% | 100,00% |
| | 853 990 | 1 431 | 855 421 | 0,17% | | |

*Table B.3.14 Events and Nonevents results and Lift rate for application propensity-to-buy CP3B model on validation sample.*

| Validation at ALL [CP3B] | | | | | | |
|---|---|---|---|---|---|---|
| Decile | Non Resp. | Resp. | Total | Resp. Rate | Cum. Resp rate | % Lift |
| ~10% | 75 661 | 343 | 76 004 | 0,45% | 0,45% | 485,85% |
| ~20% | 76 640 | 179 | 76 819 | 0,23% | 0,34% | 367,48% |
| ~30% | 76 871 | 57 | 76 928 | 0,07% | 0,25% | 271,22% |
| ~40% | 76 955 | 43 | 76 997 | 0,06% | 0,20% | 218,19% |
| ~50% | 76 985 | 29 | 77 014 | 0,04% | 0,17% | 182,42% |
| ~60% | 77 004 | 21 | 77 026 | 0,03% | 0,15% | 156,94% |
| ~70% | 77 095 | 14 | 77 109 | 0,02% | 0,13% | 137,30% |
| ~80% | 77 182 | 14 | 77 196 | 0,02% | 0,11% | 122,57% |
| ~90% | 77 187 | 7 | 77 194 | 0,01% | 0,10% | 110,01% |
| ~100% | 77 012 | 7 | 77 019 | 0,01% | 0,09% | 100,00% |
| | 768 591 | 716 | 769 307 | 0,09% | | |

In the top decile of development sample there were 84 798 customers, and there were 731 responders for a response rate of 0,86%. Compared to the average response rate of 0,17%, this gives a lift 5.15 (515%) for decile 1. In the validation sample the results are quite the same. Lift ratio gains high value of 48.6 (486%). In the top decile of validation sample there were 76 004 customers, with 343 responders and 0,45% of response rate. As average response rate is 0.09% that is the reason why Lift rate is better result than in development sample.

The next Figure presens the comparisons of Lift rates for development and validation data sets were shown.

*Figure B.3.3 Lift ratio results charts for development and validation data sets for propensity-to-buy for CP3A model.*

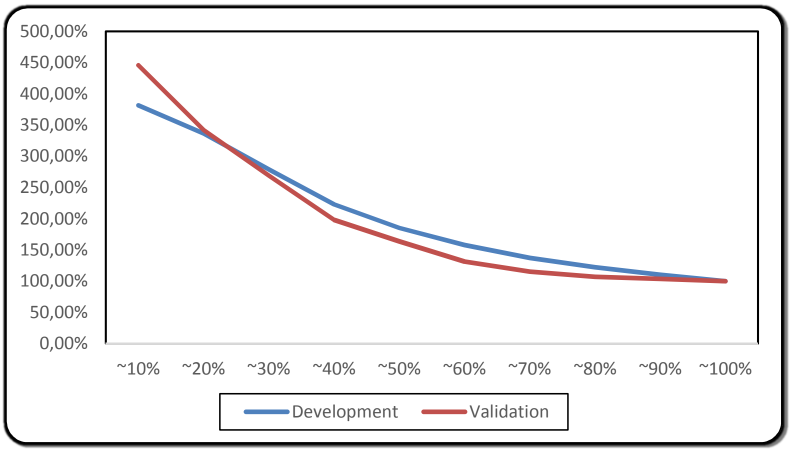
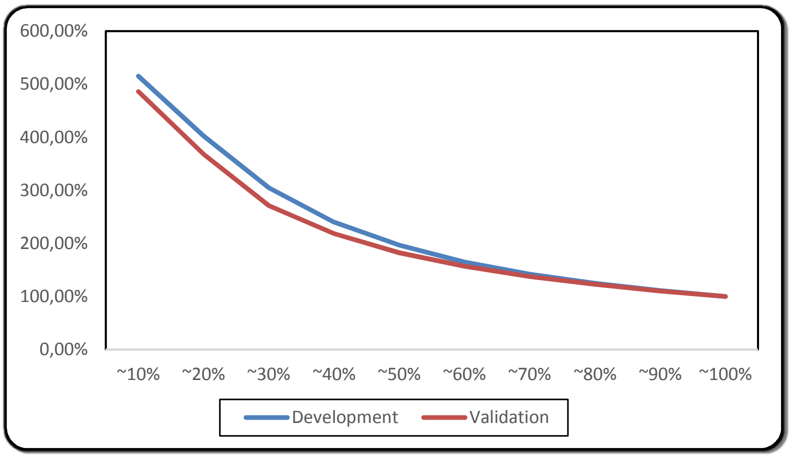



*Figure B.3.4 Lift ratio results charts for development and validation data sets for propensity-to-buy for CP3B model.*




According to the Table B.3.15 for development data set for propensity-to-buy for CP3A model the top three deciles capture 83.7% of the responders (Events). This is compared to a random baseline where three deciles (30% of the population) would capture 30% of the responders and is presented bt the Figure B.3.5.

*Table B.3.15 Event and Nonevent distribution in deciles for development sample of propensity-to-buy for CP3A model.*

| Model Development [CP3A] | | | | | | | |
|---|---|---|---|---|---|---|---|
| Approx. score % | NonResp (NR) | Responders (R) | Prob. NonResp | % of all NonResp | % of all Resp | Cum. % NonResp | Cum. % Resp |
| ~10% | 946 | 3 717 | 0,203 | 2,6% | 38,2% | 2,6% | 38,2% |
| ~20% | 1 823 | 2 840 | 0,391 | 4,9% | 29,1% | 7,5% | 67,3% |
| ~30% | 3 063 | 1 600 | 0,657 | 8,3% | 16,4% | 15,8% | 83,7% |
| ~40% | 4 136 | 528 | 0,887 | 11,2% | 5,4% | 27,0% | 89,1% |
| ~50% | 4 321 | 342 | 0,927 | 11,7% | 3,5% | 38,7% | 92,7% |
| ~60% | 4 462 | 201 | 0,957 | 12,1% | 2,1% | 50,8% | 94,7% |
| ~70% | 4 520 | 144 | 0,969 | 12,3% | 1,5% | 63,1% | 96,2% |
| ~80% | 4 504 | 159 | 0,966 | 12,2% | 1,6% | 75,3% | 97,8% |
| ~90% | 4 513 | 150 | 0,968 | 12,2% | 1,5% | 87,5% | 99,4% |
| ~100% | 4 601 | 62 | 0,987 | 12,5% | 0,6% | 100,0% | 100,0% |
| Totals | 36 889 | 9 743 | | | | | |

According to the table B.3.16 which was prepared for validation data set the top three deciles capture 61.61% of responders. Difference between these results and random baseline is illustrated also on Figure B.3.6.

*Table B.3.16 Event and Nonevent distribution in deciles for validation sample of propensity-to-buy for CP3A model.*

| Model Validation [CP3A] | | | | | | | |
|---|---|---|---|---|---|---|---|
| Approx. score % | NonResp (NR) | Responders (R) | Prob. NonResp | % of all NonResp | % of all Resp | Cum. % NonResp | Cum. % Resp |
| ~10% | 32 094 | 712 | 0,978 | 9,77% | 44,31% | 9,77% | 44,31% |
| ~20% | 18 997 | 153 | 0,992 | 5,78% | 9,52% | 15,56% | 53,83% |
| ~30% | 23 472 | 125 | 0,995 | 7,15% | 7,78% | 22,70% | 61,61% |
| ~40% | 53 462 | 258 | 0,995 | 16,28% | 16,05% | 38,98% | 77,66% |
| ~50% | 43 493 | 130 | 0,997 | 13,24% | 8,09% | 52,22% | 85,75% |
| ~60% | 53 821 | 74 | 0,998 | 16,39% | 4,60% | 68,61% | 90,35% |
| ~70% | 46 611 | 82 | 0,998 | 14,19% | 5,10% | 82,80% | 95,46% |
| ~80% | 29 545 | 47 | 0,998 | 9,00% | 2,92% | 91,80% | 98,38% |
| ~90% | 11 464 | 10 | 0,999 | 3,49% | 0,62% | 95,29% | 99,00% |
| ~100% | 15 480 | 16 | 0,999 | 4,71% | 1,00% | 100,00% | 100,00% |
| Totals | 328 438 | 1 607 | | | | | |

According to the Table B.3.17 for development data set for propensity-to-buy for CP3B model, the top two deciles capture 80.1% of the Events. It is compared to a random baseline where two deciles (20% of the population) would capture 20% of the responders and is presented by the Figure B.3.7.

*Table B.3.17 Event and Nonevent distribution in deciles for development sample of propensity-to-buy for CP3B model.*

| Model Development [CP3B] | | | | | | | |
|---|---|---|---|---|---|---|---|
| Approx. score % | NonResp (NR) | Respo nders (R) | Prob. NonResp | % of all NonResp | % of all Resp | Cum. % NonResp | Cum. % Resp |
| ~10% | 84 067 | 731 | 0,991 | 9,84% | 51,08% | 9,84% | 51,08% |
| ~20% | 85 156 | 415 | 0,995 | 9,97% | 29,00% | 19,82% | 80,08% |
| ~30% | 85 412 | 159 | 0,998 | 10,00% | 11,11% | 29,82% | 91,19% |
| ~40% | 85 505 | 66 | 0,999 | 10,01% | 4,61% | 39,83% | 95,81% |
| ~50% | 85 539 | 32 | 0,999 | 10,02% | 2,24% | 49,85% | 98,04% |
| ~60% | 85 560 | 11 | 0,999 | 10,02% | 0,77% | 59,86% | 98,81% |
| ~70% | 85 661 | 7 | 0,999 | 10,03% | 0,49% | 69,90% | 99,30% |
| ~80% | 85 757 | 7 | 0,999 | 10,04% | 0,49% | 79,94% | 99,79% |
| ~90% | 85 763 | 1 | 0,999 | 10,04% | 0,07% | 89,98% | 99,86% |
| ~100% | 85 569 | 2 | 0,999 | 10,02% | 0,14% | 100,00% | 100,00% |
| Totals | 853 990 | 1 431 | | | | | |

According to the Table B.3.18 for validation data set for propensity-to-buy for CP3B model the top two deciles capture 77.7% of the Events. It is compared to a random baseline where two deciles (20% of the population) would capture 20% of the responders and is also presented by the Figure B.3.8.

*Table B.3.18 Event and Nonevent distribution in deciles for validation sample of propensity-to- buy for CP3B model.*

| Model Validation [CP3B] | | | | | | | |
|---|---|---|---|---|---|---|---|
| Approx. score % | NonResp (NR) | Respo nders (R) | Prob. NonResp | % of all NonResp | % of all Resp | Cum. % NonResp | Cum. % Resp |
| ~10% | 75 661 | 343 | 0,995 | 9,84% | 48,00% | 9,84% | 48,00% |
| ~20% | 76 640 | 179 | 0,997 | 9,97% | 25,00% | 19,82% | 73,00% |
| ~30% | 76 871 | 57 | 0,999 | 10,00% | 8,00% | 29,82% | 81,00% |
| ~40% | 76 955 | 43 | 0,999 | 10,01% | 6,00% | 39,83% | 87,00% |
| ~50% | 76 985 | 29 | 0,999 | 10,02% | 4,00% | 49,85% | 91,00% |
| ~60% | 77 004 | 21 | 0,999 | 10,02% | 3,00% | 59,86% | 94,00% |
| ~70% | 77 095 | 14 | 0,999 | 10,03% | 2,00% | 69,90% | 96,00% |
| ~80% | 77 182 | 14 | 0,999 | 10,04% | 2,00% | 79,94% | 98,00% |
| ~90% | 77 187 | 7 | 0,999 | 10,04% | 1,00% | 89,98% | 99,00% |
| ~100% | 77 012 | 7 | 0,999 | 10,02% | 1,00% | 100,00% | 100,00% |
| Totals | 768 591 | 716 | | | | | |

The following figures illustrate the numbers discussed above.

*Figure B.3.5 Comparison of cumulative percent of Events and Nonevents for development and validation data set of propensity-to-buy for CP3A model.*

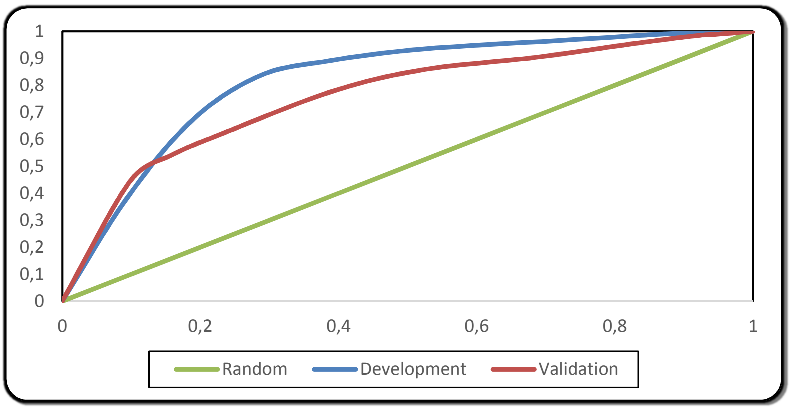
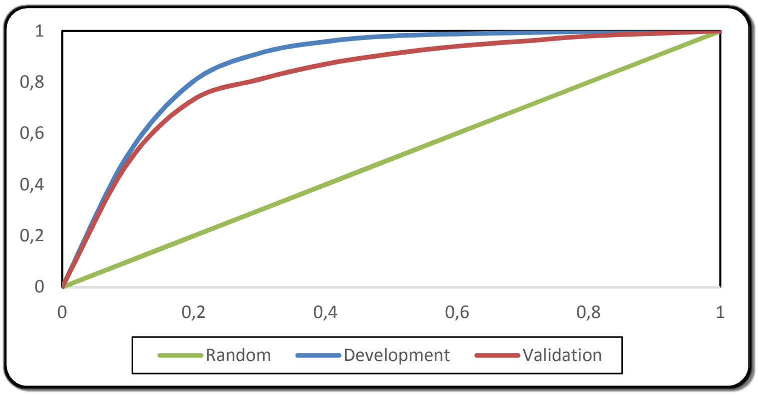



*Figure B.3.6 Comparison of cumulative percent of Events and Nonevents for development and validation data set of propensity-to-buy for CP3B model.*




The both models represent good quality of fitness to data and describe the dependence between target variable and a set of the explanatory variables.

## B.4 Propensity-to-buy Deposit Product no 1 model - details

Due to complete the picture of offered products there was a need to create one more propensity-to-buy model. Since the credit side had a big penetration of data mining support of cross-selling process it was decided to develop deposit side of bank products. As a result this model was built on the last place and it is (like propensity-to-buy model for Credit Product no 1 and Credit Product no 2) based on the historical data of previous actions regarding the behavior of customers in similar situations. Its goal is to predict how much given customer is interested in buying Deposit Product no 1 within the marketing action. There were twelve marketing campaigns which took place whole year before the time of building model with phone contact. Observation window, in the numbers included 72 390 customers who were targeted, and 14 430 who bought the DP1. It gives response rate equals 19.90%. The rule about target variable was as follows: people who increased their portfolio by buying DP1 became the target variable=1 means 'good'. The rest, who was offered and did not use it become target variable =0 means 'bad'. Response Profile table is presented below:

*Table B.4.1 Response Profile for DP1.*

| Response Profile | | | |
|---|---|---|---|
| | Ordered Value | CP1 | Total Frequency |
| | 1 | Event \| 1 | 14 430 |
| | 2 | NonEvent \| 0 | 57 987 |
| Model Convergence Status | | | |
| Convergence criterion (GCONV=1E-8) satisfied | | | |

Willingness of purchasing Deposit Product no 1 can be described with the following formula:

$$P_{DP1} = Prob(Y_{DP1}|X_{DP1}) = \frac{1}{1+e^{-L_{DP1}}} \qquad \text{(B.4.1)}$$

where

$$L_{DP1} = \alpha + \beta_1 X_{1DP1} + \beta_2 X_{2DP1} + \cdots + \beta_k X_{kDP1} \quad \text{(B.4.2)}$$

where *DP1* means Deposit Product no 1. The Table no. B.4.2 lists the parameter estimates, their standard errors and the results of the Wald test for individual parameters.

*Table B.4.2 Analysis of Maximum Likelihood Estimates for DP1.*

| Analysis of Maximum Likelihood Estimates | | | | | | | |
|---|---|---|---|---|---|---|---|
| Parameter | Parameter type | Parameter description | DF | Estimator | Standard Error | Wald Chi-Square | Pr > Chi-Sq. |
| DP1_v1 | account-related | number of active credit products no 1 which customer has got in other banks | 1 | -0.7244 | 0.0368 | 388.0253 | <.0001 |
| DP1_v2 | account-related | the number of reports of customer from credit bureau | 1 | -1.0549 | 0.0683 | 238.5508 | <.0001 |
| DP1_v3 | account-related | value of the last bought product, grouped variable | 1 | 0.0817 | 0.00658 | 153.8823 | <.0001 |
| DP1_v4 | account-related | customer has got product from group of credit products no 1 reported in other banks | 1 | -0.3304 | 0.0866 | 14.5503 | 0.0001 |
| DP1_v5 | account-related | customer had any credit product reported at other banks | 1 | -0.3682 | 0.0737 | 24.9779 | <.0001 |
| DP1_v6 | CRM-related | customer was communicated via mailing in last 24 months | 1 | -1.6801 | 0.0351 | 2286.9155 | <.0001 |
| DP1_v7 | CRM-related | customer was communicated via phone in last 3 months | 1 | 0.6199 | 0.0488 | 161.3005 | <.0001 |
| DP1_v8 | demographic | customer has got mobile phone number correct | 1 | 0.1508 | 0.00419 | 1296.8247 | <.0001 |
| DP1_v9 | demographic | size of the city that customer lives | 1 | 0.0906 | 0.00612 | 219.1440 | <.0001 |
| DP1_v10 | demographic | number of months since last bought product by customer, grouped variable | 1 | -0.3049 | 0.00842 | 1312.1666 | <.0001 |
| DP1_v11 | demographic | number of months since first bought product by customer, grouped variable | 1 | 0.0988 | 0.00767 | 165.8084 | <.0001 |
| DP1_v12 | demographic | customer gives the salary to the bank account | 1 | 1.9305 | 0.0890 | 470.5539 | <.0001 |
| DP1_v13_d1 | demographic dummy | customer bought the product from group of products no 2 as a last product | 1 | -0.8331 | 0.0518 | 258.5478 | <.0001 |
| DP1_v14 | product characteristics | sum of the credit products no 5 | 1 | 0.2020 | 0.0286 | 49.8394 | <.0001 |
| DP1_v15 | product characteristics | customer has got credit product no 8 | 1 | 0.4926 | 0.1082 | 20.7249 | <.0001 |
| DP1_v16 | product characteristics | customer has got deposit product no 2 | 1 | 1.6833 | 0.0852 | 390.2715 | <.0001 |
| DP1_v17 | product characteristics | kind of the product which was last bought by customer, grouped variable | 1 | 1.4028 | 0.0395 | 1258.4766 | <.0001 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| DP1_v18 | product characteristics | number of products in customer portfolio, grouped variable | 1 | -0.2890 | 0.0140 | 428.4006 | <.0001 |
| DP1_v19 | product characteristics | sum of the credit product no 2 | 1 | 0.5236 | 0.0319 | 268.6545 | <.0001 |
| DP1_v20 | product characteristics | sum of the credit products from group of credit products no 1 | 1 | 0.1038 | 0.0121 | 73.5910 | <.0001 |
| DP1_v21_d1 | product characteristics dummy | customer has bought deposit product no 3 as the first one | 1 | 0.8959 | 0.2122 | 17.8283 | <.0001 |
| DP1_v22_d1 | product characteristics dummy | customer has bought credit product no 3 as a last one | 1 | -0.5020 | 0.0712 | 49.7126 | <.0001 |
| DP1_v0 | | | 1 | 3.2675 | 0.2357 | 192.1391 | <.0001 |

Variables ending with 'd' means that this variable is dummy variable of the original one. Variables DP1_v8-v13 are customer-related and they describe customer from demographic point of view mostly, variables DP1_v6-v7 are linked with the CRM campaigns history and frequency and way of contact with the customer. Variables DP1_v14-v22 say about the history of the customer portfolio and DP1_v1-v5 is connected with the customer risk history.

The Association of Predicted Probabilities and Observed Responses results, which are showed in the Table B.4.3 display measures of association between predicted probabilities and observed responses which include a breakdown of the number of pairs with different responses and four rank correlation indexes. After launched the model on development sample of DP1 there were 95.1% of Concordant pairs, meaning for 95% of 8,35E+08 pairs model estimated correctly higher probability for higher ordered value than for lower ordered value. For 4.7% of total number of pairs model gave higher probability for the lower ordered value and for only 0.2% pairs probability for the both ordered values was the same.

*Table B.4.3 Association of Predicted Probabilities and Observed Responses for DP1.*

| Association of Predicted Probabilities and Observed Responses | | | |
|---|---|---|---|
| Percent Concordant | 95.1 | Somers' D | 0.903 |
| Percent Discordant | 4.7 | Gamma | 0.905 |
| Percent Tied | 0.2 | Tau-a | 0.288 |
| Pairs | 835186761 | c | 0.952 |

Odds ratios estimates for parameters used in propensity-to-buy for DP1 model are able to find below.

*Table B.4.4 Odds ratio estimates for DP1.*

| Odds Ratio Estimates | | | | |
|---|---|---|---|---|
| Effect | Parameter description | Point Estimate | 95% Wald Confidence Limits | |
| DP1_v1 | number of active credit products no 1 which customer has got in other banks | 0.485 | 0.451 | 0.521 |
| DP1_v2 | the number of reports of customer from credit bureau | 0.348 | 0.305 | 0.398 |
| DP1_v3 | value of the last bought product, grouped variable | 1.085 | 1.071 | 1.099 |
| DP1_v4 | customer has got product from group of credit products no 1 reported in other banks | 0.719 | 0.606 | 0.852 |
| DP1_v5 | customer had any credit product reported at other banks | 0.692 | 0.599 | 0.799 |
| DP1_v6 | customer was communicated via mailing in last 24 months | 0.186 | 0.174 | 0.200 |
| DP1_v7 | customer was communicated via phone in last 3 months | 1.859 | 1.689 | 2.045 |
| DP1_v8 | customer has got mobile phone number correct | 1.163 | 1.153 | 1.172 |
| DP1_v9 | size of the city that customer lives | 1.095 | 1.082 | 1.108 |
| DP1_v10 | number of months since last bought product by customer, grouped variable | 0.737 | 0.725 | 0.749 |
| DP1_v11 | number of months since first bought product by customer, grouped variable | 1.104 | 1.087 | 1.121 |
| DP1_v12 | customer gives the salary to the bank account | 6.893 | 5.790 | 8.206 |
| DP1_v13_d1 | customer bought the product from group of products no 2 as a last product | 0.435 | 0.393 | 0.481 |
| DP1_v14 | sum of the credit products no 5 | 1.224 | 1.157 | 1.294 |
| DP1_v15 | customer has got credit product no 8 | 1.636 | 1.324 | 2.023 |
| DP1_v16 | customer has got deposit product no 2 | 5.383 | 4.555 | 6.362 |

| | | | | |
|---|---|---|---|---|
| DP1_v17 | kind of the product which was last bought by customer, grouped variable | 4.067 | 3.763 | 4.394 |
| DP1_v18 | number of products in customer portfolio, grouped variable | 0.749 | 0.729 | 0.770 |
| DP1_v19 | sum of the credit product no 2 | 1.688 | 1.586 | 1.797 |
| DP1_v20 | sum of the credit products from group of credit products no 1 | 1.109 | 1.083 | 1.136 |
| DP1_v21_d1 | customer has bought deposit product no 3 as the first one | 2.450 | 1.616 | 3.713 |
| DP1_v22_d1 | customer has bought credit product no 3 as a last one | 0.605 | 0.527 | 0.696 |

For variable DP1_v8, where odds ratio=1.163 it can be interpreted as follows: if there are the customers who are different from each other only in terms of having correct mobile phone number, the customer, who has an odds ratio of buying DP1 versus not buying DP1 which is 16.3% higher than the customer whose mobile phone number is not valid. For variable DP1_v2, where odds ratio=0.348 it can be in turn described in a way: if there are customers who are different from each other only in terms of reports in credit bureau, the customer who has an odds ratio of buying DP1 versus not buying DP1 which is 65.2% lower than the customer whose got no report or steady number of reports in credit bureau.

The Table B.3.4.5 illustrate the Hosmer and Lemeshow test results of propensity-to-buy DP1 model, which examined model fit to the observed values. As for previous models, all observations were divided into 10 groups according to the increasing probability, which examined the distribution of values compatible with the observed distribution of the theoretical value. The null hypothesis implies a good fit to the data model against the alternative one meaning a bad match. In this model there is no reason to reject the null hypothesis, what can allow to conclude that developed model is well suited to the data.

*Table B.4.5 The Hosmer-Lemeshow goodness-of-fit results of propensity-to-buy DP1 model.*

| Hosmer and Lemeshow Goodness-of-Fit Test | | | | | |
|---|---|---|---|---|---|
| Chi-square | | | DF | Pr>Chi-square | |
| 4.9550 | | | 8 | 0.6322 | |
| Hosmer-Lemeshow test results | | | | | |
| Group | Total | Target_variable=1 | | Target_variable=0 | |
| | | Observed | Predicted | Observed | Predicted |
| 1 | 7 239 | 53 | 39.40 | 7 186 | 7199.60 |
| 2 | 7 240 | 92 | 84.13 | 7 148 | 7155.87 |
| 3 | 7 239 | 94 | 124.46 | 7 145 | 7114.54 |
| 4 | 7 241 | 124 | 169.41 | 7117 | 7071.59 |
| 5 | 7 239 | 214 | 231.69 | 7025 | 7007.31 |
| 6 | 7 240 | 277 | 337.87 | 6963 | 6902.13 |
| 7 | 7 240 | 492 | 554.44 | 6748 | 6685.56 |
| 8 | 7 239 | 1 137 | 1081.26 | 6102 | 6157.74 |
| 9 | 7 239 | 4 802 | 4602.83 | 2437 | 2636.17 |
| 10 | 7 234 | 7 118 | 7177.57 | 116 | 56.43 |

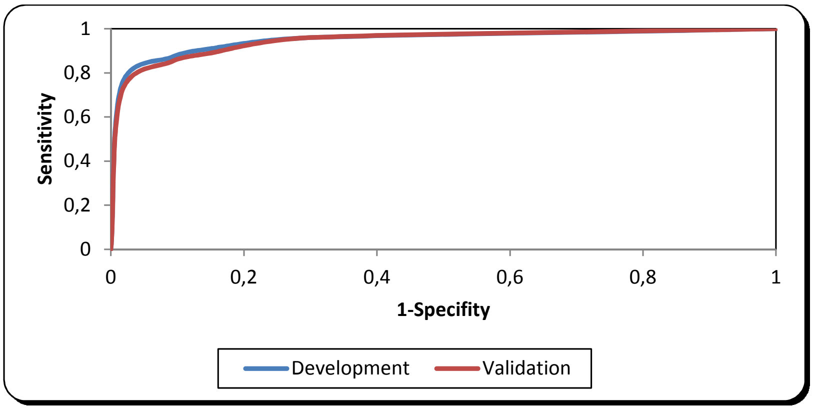
Table B.4.6 displays the Classification results of created propensity-to-buy for Deposit Product no 1 model.

*Table B.4.6 Classification table results of propensity-to-buy DP1 model.*

| Classification table | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Prob | Correct | | Incorrect | | Percentage | | | | |
| Level | Event | Nonevent | Event | Nonevent | Correct | Sensitivity | Specificity | False POS | False NEG |
| 0.500 | 10720 | 57064 | 3683 | 923 | 93.6 | 74.4 | 98.4 | 7.9 | 6.1 |

Developed model has represented correctness of 93.6% of observations on the 0.5 probability level. When target variable is 1 there are 10 720 observations which are corrected assigned and for target variable equals 0 there are 57 064 correct assignments. Moreover, for 7.9% of observations model gave the 0 value while it should be 1 value and for two percentage less, 6.1% it gave the 1 value while it should be 0 value. Figure B.4.1 shows ROC curves. These curves show model fit for development and validation samples, based on Sensitivity and Specificity values.

*Figure B.4.1 ROC charts for development and validation data sets of propensity-to-buy DP1 model.*




The ROC curves confirm the assumption that created model fits the observations and it is going to predict marketing offer responders correctly. Table B.4.7 shows the number of Events and Nonevents and Lift rate value for each decile of Development sample and Table B.4.8 shows the same indicators but after applying the propensity-to-buy for DP1 model on the validation data set.

*Table B.4.7 Events and Nonevents results and Lift rate for application propensity-to-buy DP1 model on development sample.*

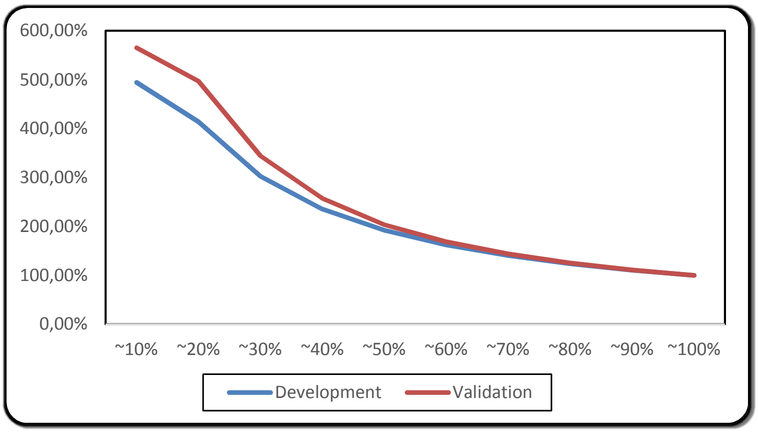
| Development at ALL | | | | | | |
|---|---|---|---|---|---|---|
| Decile | Non Resp. | Resp. | Total | Resp. Rate | Cum. Resp rate | % Lift |
| ~10% | 116 | 7 118 | 7 234 | 98,40% | 98,40% | 494,54% |
| ~20% | 2437 | 4 802 | 7 239 | 66,34% | 82,36% | 413,95% |
| ~30% | 6102 | 1 137 | 7 239 | 15,71% | 60,14% | 302,25% |
| ~40% | 6748 | 492 | 7 240 | 6,80% | 46,80% | 235,21% |
| ~50% | 6963 | 277 | 7 240 | 3,83% | 38,20% | 192,00% |
| ~60% | 7025 | 214 | 7 239 | 2,96% | 32,33% | 162,48% |
| ~70% | 7117 | 124 | 7 241 | 1,71% | 27,95% | 140,49% |
| ~80% | 7 145 | 94 | 7 239 | 1,30% | 24,62% | 123,74% |
| ~90% | 7 148 | 92 | 7 240 | 1,27% | 22,03% | 110,70% |
| ~100% | 7 186 | 53 | 7 239 | 0,73% | 19,90% | 100,00% |
| | 57 987 | 14 403 | 72 390 | 19,90% | | |

*Table B.4.8 Events and Nonevents results and Lift rate for application propensity-to-buy DP1 model on validation sample.*

| Validation at ALL | | | | | | |
|---|---|---|---|---|---|---|
| Decile | Non Resp. | Resp. | Total | Resp. Rate | Cum. Resp rate | % Lift |
| ~10% | 12 | 1 597 | 1 609 | 99,25% | 99,25% | 565,04% |
| ~20% | 368 | 1 021 | 1 389 | 73,51% | 87,32% | 497,13% |
| ~30% | 1521 | 289 | 1 810 | 15,97% | 60,46% | 344,20% |
| ~40% | 1724 | 86 | 1 810 | 4,75% | 45,23% | 257,46% |
| ~50% | 1790 | 20 | 1 810 | 1,10% | 35,75% | 203,52% |
| ~60% | 1794 | 19 | 1 813 | 1,05% | 29,61% | 168,54% |
| ~70% | 1799 | 8 | 1 807 | 0,44% | 25,23% | 143,64% |
| ~80% | 1 798 | 12 | 1 810 | 0,66% | 22,02% | 125,38% |
| ~90% | 1 798 | 12 | 1 810 | 0,66% | 19,56% | 111,33% |
| ~100% | 1 803 | 6 | 1 809 | 0,33% | 17,57% | 100,00% |
| | 14 407 | 3 070 | 17 477 | 17,57% | | |

In the top decile of development sample there were 7 234 customers, with 7 118 responders for a response rate of 98.40%. Compared to the average response rate of 19.90%, this gives a lift 4.95 (495%) for decile 1. In the validation sample the results are similar, even though the numbers of validation sample are different. In the top decile of validation sample there were 1 809 customers, with 1 797 responders and 99.34% of response rate. As average response rate is 20.39% Lift rate equals 4.87 (487%) already, what is a bit worse result than in dDevelopment sample. The Figure B.4.2 shows the comparison of Lift rates for development and validation data sets.

*Figure B.4.2 Lift ratio results charts for development and validation data sets for propensity-to-buy for DP1 model.*



Next Figure B.4.5 compares the cumulative percent of responses captured as each decile is added to the target.

According to the Table B.4.9 for development data set the top two deciles capture 88.0% of the responders. This is compared to a random baseline where two deciles (20% of the population) would capture 20% of the responders. The greater the area between two lines – baseline and line with the information of cumulative percent of Responses Captured, the more the model is able to concentrate responders in the top deciles.

*Table B.4.9 Event and Nonevent distribution in deciles for development sample of propensity-to-buy for DP1 model.*

| Development at ALL | | | | | | | |
|---|---|---|---|---|---|---|---|
| Approx. score % | NonResp (NR) | Responders (R) | Prob. NonResp | % of all NonResp | % of all Resp | Cum. % NonResp | Cum. % Resp |
| ~10% | 37 | 7 202 | 0,005 | 0,1% | 50,0% | 0,1% | 50,0% |
| ~20% | 1 770 | 5 469 | 0,245 | 3,1% | 38,0% | 3,1% | 88,0% |
| ~30% | 6 214 | 1 024 | 0,859 | 10,7% | 7,1% | 13,8% | 95,1% |
| ~40% | 6 895 | 345 | 0,952 | 11,9% | 2,4% | 25,7% | 97,5% |
| ~50% | 7 109 | 129 | 0,982 | 12,3% | 0,9% | 38,0% | 98,4% |
| ~60% | 7 162 | 78 | 0,989 | 12,4% | 0,5% | 50,3% | 98,9% |
| ~70% | 7 194 | 47 | 0,994 | 12,4% | 0,3% | 62,7% | 99,2% |
| ~80% | 7 192 | 43 | 0,994 | 12,4% | 0,3% | 75,1% | 99,5% |
| ~90% | 7 201 | 40 | 0,994 | 12,4% | 0,3% | 87,6% | 99,8% |
| ~100% | 7 213 | 26 | 0,996 | 12,4% | 0,2% | 100,0% | 100,0% |
| Totals | 57 987 | 14 403 | | | | | |

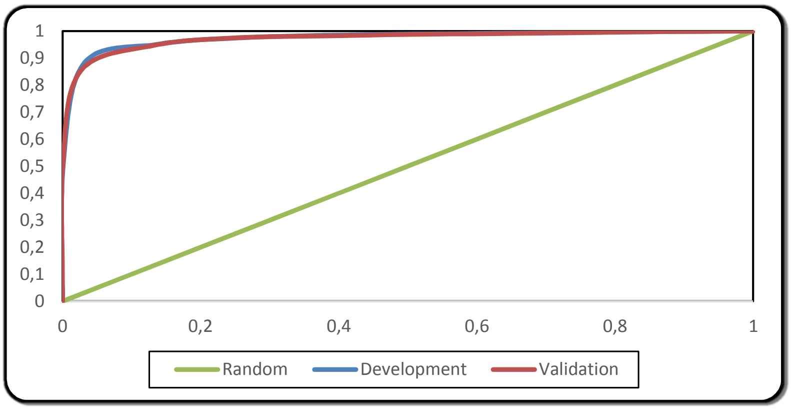
According to the table B.4.9 prepared for validation data set the top two deciles capture 85.3% of responder, instead of 20% according to the baseline.

*Table B.4.10 Event and Nonevent distribution in deciles for validation sample of propensity-to-buy for DP1 model*.

| Validation at ALL | | | | | | | |
|---|---|---|---|---|---|---|---|
| Approx. score % | NonResp (NR) | Respond ers | Prob. NonResp | % of all NonResp | % of all Resp | Cum. % NonResp | Cum. % Resp |
| ~10% | 12 | 1 597 | 0,007 | 0,08% | 52,02% | 0,08% | 52,02% |
| ~20% | 368 | 1 021 | 0,265 | 2,55% | 33,26% | 2,64% | 85,28% |
| ~30% | 1 521 | 289 | 0,840 | 10,56% | 9,41% | 13,19% | 94,69% |
| ~40% | 1 724 | 86 | 0,952 | 11,97% | 2,80% | 25,16% | 97,49% |
| ~50% | 1 790 | 20 | 0,989 | 12,42% | 0,65% | 37,59% | 98,14% |
| ~60% | 1 794 | 19 | 0,990 | 12,45% | 0,62% | 50,04% | 98,76% |
| ~70% | 1 799 | 8 | 0,996 | 12,49% | 0,26% | 62,53% | 99,02% |
| ~80% | 1 798 | 12 | 0,993 | 12,48% | 0,39% | 75,01% | 99,41% |
| ~90% | 1 798 | 12 | 0,993 | 12,48% | 0,39% | 87,49% | 99,80% |
| ~100% | 1 803 | 6 | 0,997 | 12,51% | 0,20% | 100,00% | 100,00% |
| Totals | 14 407 | 3 070 | | | | | |

The following Figure B.4.3 illustrates the numbers discussed above and presented in the tables above.

*Figure B.4.3 Comparison of cumulative percent of Events and Nonevents for development and validation data set of propensity-to-buy for DP1 model.*



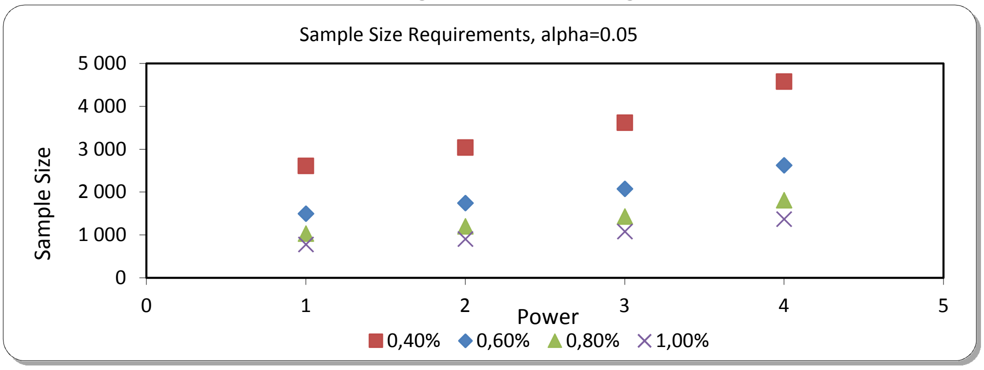

Both curves are above the random baseline with large distance between. Model, like four previous ones, represents good fitness to data

# C. Minimal sample size defining

This Appendix refers to Chapter 5.1 describing proposed MPO test design and presents the Charts associated with every proposed Segment. These charts are the results of the exercises of finding the most optimal minimal sample size for Control and Test groups.

Figure C.1 shows results for Segment 1 and Segment 4 (minimum response rates in these segments are almost the same).

*Figure C.1 Dependence between Power and Sample Size for several Effect sizes and two values of α for Segment 1 and Segment 4.*



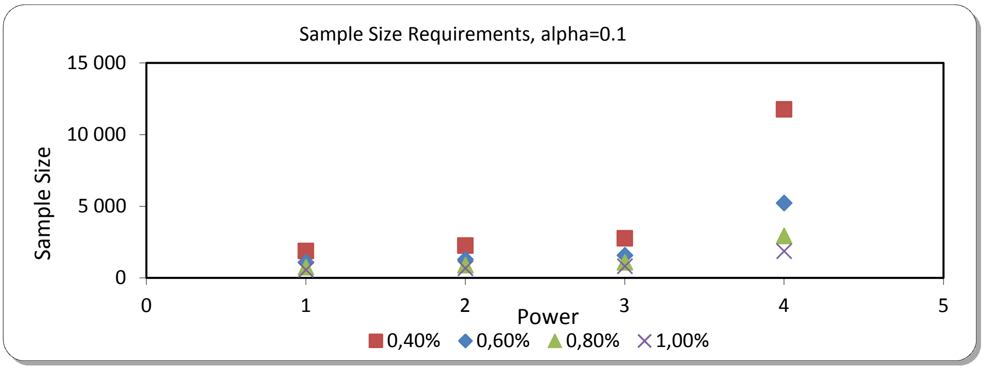
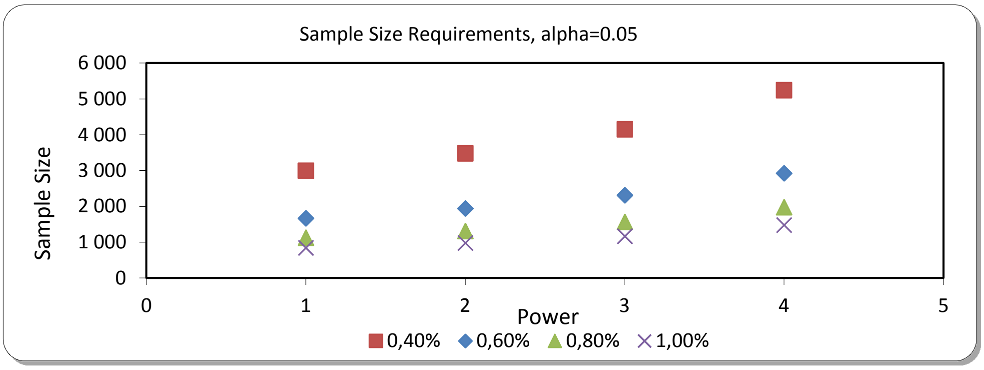



Figure C.2 shows results of plot dependences between Power and Sample Size for Segment 2.

*Figure C.2 Dependence between Power and Sample Size for several Effect sizes and two values of α for Segment 2.*



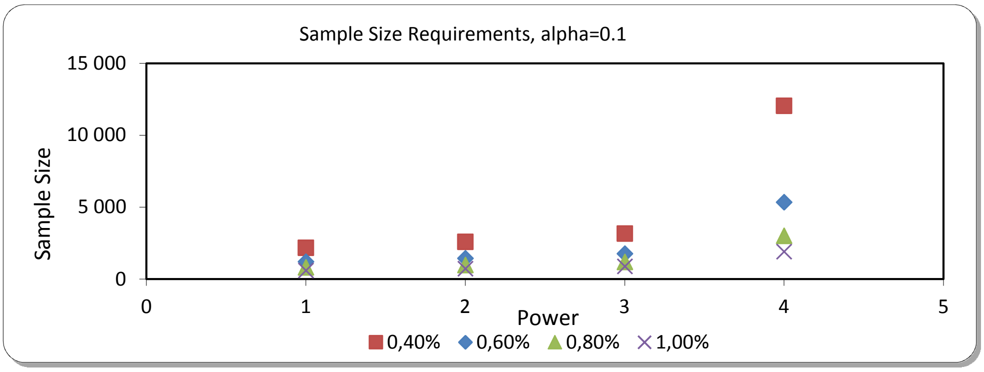
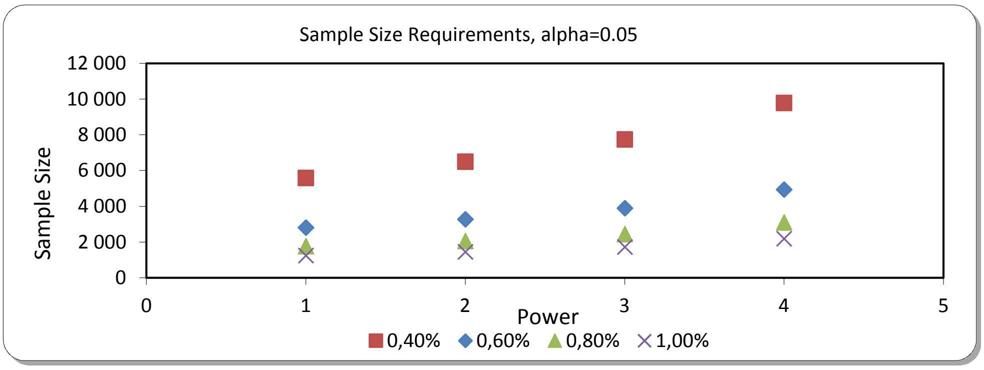



Figure C.3 presents results of a plot dependences between Power and Sample Size for Segment 3.

*Figure C.3 Dependence between Power and Sample Size for several Effect sizes and two values of α for Segment 3.*



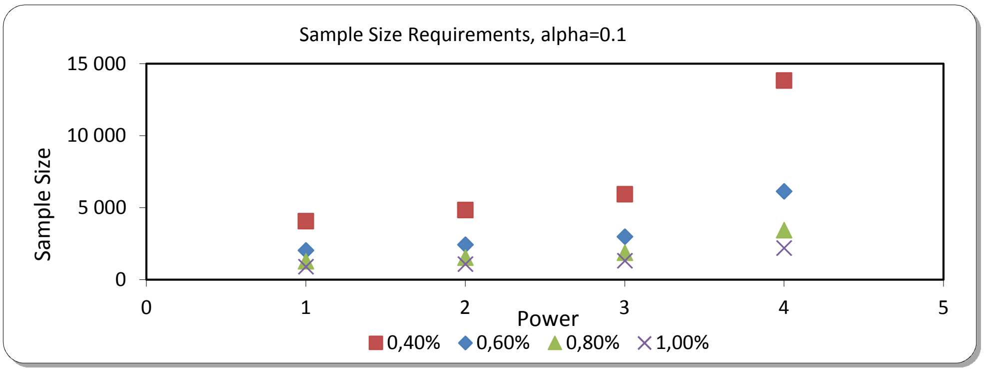
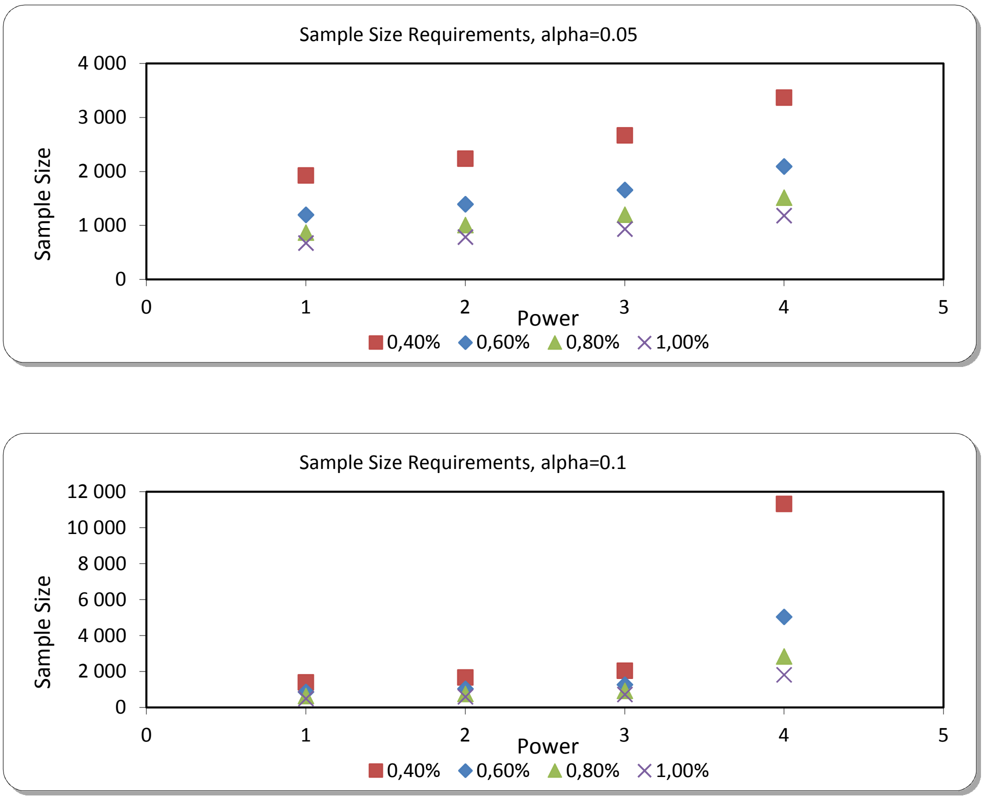



Figure C.4 represents results of a plot dependences between Power and Sample Size for Segment 5 are presented.

*Figure C.4 Dependence between Power and Sample Size for several Effect sizes and two values of α for Segment 5.*



## D. MPO Test Results

In this part there are original results of MPO test presented. They are the source of all the calculations and results from Chapter 5. The following table presents information about the volumes which were bought with a concrete product by a given customer in the particular groups and segments. Sales are connected with all the purchased products regardless of the received offer. Based on the response rates expressed in number of customers, value of average ticket in given group is possible to be counted. All the values of volumes and average tickets should be multiple by one thousand to become a real values achieved in the MPO test, but they are still expressed in currency units adopted specifically for this research. The results for every group and particular bank products are shown from the segments perspective. Moreover, each segment is summarized by summing up the bought products and volumes, by counting the response rates and average tickets but with keeping the division into the sort of the product.

*Table D.1 Summary of response rates and volumes achieved in the MPO test in particular groups, segments and in division on bank products, regardless of the received offer.*

| | | CP1 | | | | CP2 | | | | CP3 | | | | CP4 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TEST GROUP | OFFER TYPE | RR NO | RR % | VOLUME [CU][9] | AVG TICKET [CU] | RR NO | RR % | VOLUME [CU] | AVG TICKET [CU] | RR NO | RR % | VOLUME [CU] | AVG TICKET [CU] | RR NO | RR % | VOLUME [CU] | AVG TICKET [CU] |
| SEG_1_CL_01 | No offer | 135 | 2,08% | 3 326 | 25 | 32 | 0,48% | 209 | 7 | 27 | 0,42% | 36 | 1 | 32 | 0,48% | 92 | 3 |
| SEG_1_CL_02 | CP2 | 194 | 2,98% | 4 757 | 25 | 54 | 0,83% | 119 | 2 | 18 | 0,28% | 40 | 2 | 36 | 0,55% | 258 | 7 |
| SEG_1_CL_03 | DP1 | 113 | 1,73% | 1 885 | 17 | 36 | 0,55% | 86 | 2 | 5 | 0,07% | 11 | 2 | 41 | 0,62% | 292 | 7 |
| SEG_1_CL_04 | CP3 | 131 | 2,01% | 2 644 | 20 | 14 | 0,21% | 37 | 3 | 23 | 0,35% | 38 | 2 | 32 | 0,48% | 233 | 7 |
| SEG_1_CL_05 | CP1\|CP2\|CP3 | 212 | 3,25% | 4 486 | 21 | 59 | 0,90% | 178 | 3 | 18 | 0,28% | 26 | 1 | 36 | 0,55% | 256 | 7 |
| SEG_1_CL_06 | CP1\|CP2\|DP1 | 131 | 2,01% | 3 594 | 28 | 36 | 0,55% | 82 | 2 | 14 | 0,21% | 28 | 2 | 27 | 0,42% | 122 | 5 |
| SEG_1_CL_07 | CP1 | 225 | 3,46% | 6 240 | 28 | 32 | 0,48% | 95 | 3 | 23 | 0,35% | 29 | 1 | 54 | 0,83% | 240 | 4 |
| SEG_1_CM_00 | CP1\|CP2 | 590 | 3,12% | 9 537 | 16 | 212 | 1,12% | 586 | 3 | 81 | 0,43% | 149 | 2 | 140 | 0,74% | 2 188 | 16 |
| *SEG_1_sum* | | *1 728* | *2,68%* | *36 468* | *21* | *473* | *0,73%* | *1 392* | *3* | *207* | *0,32%* | *357* | *2* | *396* | *0,61%* | *3 682* | *9* |
| SEG_2_CL_01 | CP1 | 250 | 3,85% | 5 736 | 23 | 14 | 0,22% | 98 | 7 | 43 | 0,66% | 52 | 1 | 172 | 2,64% | 2 005 | 12 |
| SEG_2_CL_02 | CP2 | 261 | 4,01% | 4 922 | 19 | 151 | 2,32% | 229 | 2 | 41 | 0,63% | 55 | 1 | 343 | 5,28% | 7 284 | 21 |
| SEG_2_CL_03 | DP1 | 94 | 1,44% | 1 931 | 21 | 10 | 0,16% | 14 | 1 | 31 | 0,48% | 78 | 3 | 114 | 1,76% | 1 974 | 17 |
| SEG_2_CL_04 | CP3 | 173 | 2,66% | 2 423 | 14 | 106 | 1,63% | 159 | 2 | 106 | 1,63% | 275 | 3 | 159 | 2,45% | 1 756 | 11 |
| SEG_2_CL_05 | CP1\|CP2\|CP3 | 202 | 3,11% | 4 074 | 20 | 83 | 1,28% | 109 | 1 | 48 | 0,73% | 68 | 1 | 190 | 2,92% | 2 577 | 14 |
| SEG_2_CL_06 | No offer | 257 | 3,95% | 6 466 | 25 | 110 | 1,69% | 191 | 2 | 49 | 0,75% | 217 | 4 | 135 | 2,07% | 3 455 | 26 |
| SEG_2_CL_07 | CP1\|CP2 | 251 | 3,86% | 3 665 | 15 | 155 | 2,39% | 368 | 2 | 121 | 1,86% | 294 | 2 | 50 | 0,76% | 5 938 | 120 |
| SEG_2_CM_00 | CP1\|CP2\|DP1 | 525 | 2,78% | 9 351 | 18 | 125 | 0,66% | 167 | 1 | 200 | 1,06% | 447 | 2 | 626 | 3,31% | 11 735 | 19 |

[9] [CU] Current Unit

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *SEG_2_sum* | | *2 012* | *3,12%* | *38 568* | *19* | *756* | *1,17%* | *1 335* | *2* | *639* | *0,99%* | *1 486* | *2* | *1 788* | *2,78%* | *36 724* | *21* |
| SEG_3_CL_01 | CP1 | 28 | 0,43% | 769 | 27 | 0 | 0,00% | 0 | 0 | 0 | 0,00% | 0 | 0 | 14 | 0,22% | 257 | 18 |
| SEG_3_CL_02 | CP2 | 21 | 0,32% | 776 | 37 | 21 | 0,32% | 144 | 7 | 7 | 0,11% | 23 | 3 | 7 | 0,11% | 187 | 27 |
| SEG_3_CL_03 | No offer | 18 | 0,28% | 209 | 12 | 5 | 0,07% | 6 | 1 | 9 | 0,14% | 9 | 1 | 9 | 0,14% | 204 | 23 |
| SEG_3_CL_04 | CP3 | 28 | 0,43% | 888 | 32 | 0 | 0,00% | 0 | 0 | 0 | 0,00% | 0 | 0 | 0 | 0,00% | 0 | 0 |
| SEG_3_CL_05 | CP1\|CP2\|CP3 | 28 | 0,43% | 1 179 | 43 | 14 | 0,21% | 37 | 3 | 0 | 0,00% | 0 | 0 | 7 | 0,11% | 91 | 13 |
| SEG_3_CL_06 | CP1\|CP2\|DP1 | 33 | 0,51% | 1 682 | 51 | 20 | 0,31% | 74 | 4 | 7 | 0,10% | 4 | 1 | 0 | 0,00% | 0 | 0 |
| SEG_3_CL_07 | DP1\|CP3 | 0 | 0,00% | 0 | 0 | 0 | 0,00% | 0 | 0 | 0 | 0,00% | 0 | 0 | 10 | 0,16% | 77 | 7 |
| SEG_3_CL_08 | CP1\|CP2 | 35 | 0,54% | 394 | 11 | 10 | 0,15% | 27 | 3 | 5 | 0,07% | 6 | 1 | 5 | 0,07% | 89 | 20 |
| SEG_3_CM_00 | DP1 | 40 | 0,21% | 1 767 | 44 | 40 | 0,21% | 320 | 8 | 0 | 0,00% | 0 | 0 | 30 | 0,16% | 152 | 5 |
| *SEG_3_sum* | | *231* | *0,33%* | *7 663* | *33* | *109* | *0,15%* | *607* | *6* | *27* | *0,04%* | *43* | *2* | *82* | *0,12%* | *1 057* | *13* |
| SEG_4_CL_01 | CP1 | 377 | 5,80% | 10 653 | 28 | 78 | 1,19% | 189 | 2 | 67 | 1,02% | 116 | 2 | 144 | 2,22% | 2 845 | 20 |
| SEG_4_CL_02 | CP2 | 341 | 5,24% | 11 714 | 34 | 76 | 1,17% | 197 | 3 | 47 | 0,72% | 102 | 2 | 123 | 1,90% | 2 530 | 21 |
| SEG_4_CL_03 | DP1 | 277 | 4,26% | 7 807 | 28 | 96 | 1,48% | 203 | 2 | 51 | 0,78% | 94 | 2 | 90 | 1,39% | 1 094 | 12 |
| SEG_4_CL_04 | CP3 | 284 | 4,37% | 9 679 | 34 | 63 | 0,96% | 294 | 5 | 51 | 0,79% | 114 | 2 | 131 | 2,01% | 972 | 7 |
| SEG_4_CL_05 | No offer | 172 | 2,65% | 3 896 | 23 | 65 | 0,99% | 180 | 3 | 43 | 0,66% | 86 | 2 | 97 | 1,49% | 1 717 | 18 |
| SEG_4_CL_06 | CP1\|CP2\|DP1 | 179 | 2,76% | 4 657 | 26 | 56 | 0,87% | 127 | 2 | 39 | 0,59% | 106 | 3 | 70 | 1,08% | 1 456 | 21 |
| SEG_4_CL_07 | CP1\|CP2 | 362 | 5,57% | 6 863 | 19 | 104 | 1,60% | 284 | 3 | 66 | 1,02% | 137 | 2 | 90 | 1,38% | 1 347 | 15 |
| SEG_4_CM_00 | CP1\|CP2\|CP3 | 1 055 | 5,58% | 35 063 | 33 | 354 | 1,88% | 892 | 3 | 190 | 1,00% | 297 | 2 | 297 | 1,57% | 4 428 | 15 |
| *SEG_4_sum* | | *3 278* | *4,62%* | *97 997* | *30* | *1 000* | *1,41%* | *2 973* | *3* | *581* | *0,82%* | *1 096* | *2* | *1 125* | *1,59%* | *17 446* | *16* |
| SEG_5_CL_01 | CP1 | 34 | 0,52% | 552 | 16 | 3 | 0,05% | 4 | 1 | 6 | 0,09% | 6 | 1 | 15 | 0,24% | 415 | 27 |
| SEG_5_CL_02 | CP2 | 14 | 0,22% | 200 | 14 | 14 | 0,22% | 79 | 5 | 6 | 0,09% | 12 | 2 | 6 | 0,09% | 28 | 5 |
| SEG_5_CL_03 | DP1 | 16 | 0,24% | 693 | 44 | 2 | 0,03% | 27 | 15 | 0 | 0,00% | 0 | 0 | 10 | 0,16% | 41 | 4 |
| SEG_5_CL_04 | CP3 | 18 | 0,27% | 310 | 17 | 3 | 0,05% | 4 | 1 | 0 | 0,00% | 0 | 0 | 6 | 0,09% | 29 | 5 |
| SEG_5_CL_05 | CP1\|CP2\|CP3 | 9 | 0,13% | 174 | 20 | 6 | 0,09% | 20 | 3 | 0 | 0,00% | 0 | 0 | 12 | 0,18% | 85 | 7 |
| SEG_5_CL_06 | CP1\|CP2\|DP1 | 14 | 0,22% | 1 004 | 69 | 6 | 0,09% | 8 | 1 | 3 | 0,04% | 10 | 3 | 12 | 0,18% | 127 | 11 |
| SEG_5_CL_07 | DP1\|CP3 | 12 | 0,19% | 513 | 41 | 12 | 0,19% | 17 | 1 | 0 | 0,00% | 0 | 0 | 6 | 0,10% | 0 | 0 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SEG_5_CL_08 | CP1\|CP2 | 26 | 0,40% | 396 | 15 | 23 | 0,35% | 93 | 4 | 5 | 0,08% | 11 | 2 | 5 | 0,07% | 103 | 23 |
| SEG_5_CM_00 | No offer | 35 | 0,18% | 780 | 23 | 10 | 0,05% | 39 | 4 | 0 | 0,00% | 0 | 0 | 5 | 0,03% | 0 | 0 |
| *SEG_5_sum* | | *178* | *0,25%* | *4 621* | *26* | *79* | *0,11%* | *291* | *4* | *20* | *0,03%* | *39* | *2* | *76* | *0,11%* | *829* | *11* |

Because the first, profit side of the financial overview of MPO test has been just described, the cost side should be also briefly explained. Since the MPO test assumes, that every control group consists of 6 500 customers and every test group includes 18 900 customers, the costs of all control groups are the same and the costs of all the tests groups are the same as well.

*Table D.2 The financial overview of the Segments.*

| TEST GROUP | OFFER TYPE | TOTAL VOLUME [CU] | INCOME [CU] | COSTS [CU] | NET INCOME [CU] |
|---|---|---|---|---|---|
| SEG_1_CL_01 | No offer | 3 663 | 247 | 10 | 237 |
| SEG_1_CL_02 | CP2 | 5 174 | 340 | 10 | 330 |
| SEG_1_CL_03 | DP1 | 2 274 | 138 | 10 | 128 |
| SEG_1_CL_04 | CP3 | 2 953 | 188 | 10 | 179 |
| SEG_1_CL_05 | CP1\|CP2\|CP3 | 4 945 | 324 | 10 | 315 |
| SEG_1_CL_06 | CP1\|CP2\|DP1 | 3 827 | 256 | 10 | 246 |
| SEG_1_CL_07 | CP1 | 6 602 | 440 | 10 | 430 |
| SEG_1_CM_00 | CP1\|CP2 | 12 460 | 716 | 28 | 688 |
| *SEG_1_sum* | | *41 898* | *2 650* | *96* | *2 553* |
| SEG_2_CL_01 | CP1 | 7 892 | 413 | 10 | 403 |
| SEG_2_CL_02 | CP2 | 12 490 | 384 | 10 | 375 |
| SEG_2_CL_03 | DP1 | 3 998 | 146 | 10 | 137 |
| SEG_2_CL_04 | CP3 | 4 613 | 203 | 10 | 194 |
| SEG_2_CL_05 | CP1\|CP2\|CP3 | 6 829 | 302 | 10 | 293 |
| SEG_2_CL_06 | No offer | 10 329 | 486 | 10 | 477 |
| SEG_2_CL_07 | CP1\|CP2 | 10 265 | 319 | 10 | 309 |
| SEG_2_CM_00 | CP1\|CP2\|DP1 | 21 700 | 728 | 28 | 700 |
| *SEG_2_sum* | | *78 114* | *2 983* | *96* | *2 886* |
| SEG_3_CL_01 | CP1 | 1 025 | 54 | 10 | 44 |
| SEG_3_CL_02 | CP2 | 1 130 | 66 | 10 | 56 |
| SEG_3_CL_03 | No offer | 429 | 16 | 10 | 6 |
| SEG_3_CL_04 | CP3 | 888 | 61 | 10 | 52 |
| SEG_3_CL_05 | CP1\|CP2\|CP3 | 1 307 | 84 | 10 | 74 |
| SEG_3_CL_06 | CP1\|CP2\|DP1 | 1 760 | 121 | 10 | 112 |
| SEG_3_CL_07 | DP1\|CP3 | 77 | 0 | 10 | -9 |
| SEG_3_CL_08 | CP1\|CP2 | 516 | 30 | 10 | 20 |
| SEG_3_CM_00 | DP1 | 2 239 | 145 | 28 | 116 |
| *SEG_3_sum* | | *9 370* | *577* | *106* | *471* |
| SEG_4_CL_01 | CP1 | 13 803 | 766 | 10 | 756 |
| SEG_4_CL_02 | CP2 | 14 543 | 838 | 10 | 828 |
| SEG_4_CL_03 | DP1 | 9 199 | 563 | 10 | 553 |
| SEG_4_CL_04 | CP3 | 11 060 | 699 | 10 | 690 |
| SEG_4_CL_05 | No offer | 5 879 | 293 | 10 | 283 |
| SEG_4_CL_06 | CP1\|CP2\|DP1 | 6 346 | 342 | 10 | 333 |
| SEG_4_CL_07 | CP1\|CP2 | 8 632 | 507 | 10 | 498 |
| SEG_4_CM_00 | CP1\|CP2\|CP3 | 40 681 | 2 517 | 28 | 2 488 |
| *SEG_4_sum* | | *119 512* | *7 103* | *96* | *7 007* |
| SEG_5_CL_01 | CP1 | 977 | 40 | 10 | 30 |
| SEG_5_CL_02 | CP2 | 319 | 20 | 10 | 11 |
| SEG_5_CL_03 | DP1 | 761 | 50 | 10 | 40 |
| SEG_5_CL_04 | CP3 | 343 | 22 | 10 | 12 |
| SEG_5_CL_05 | CP1\|CP2\|CP3 | 278 | 14 | 10 | 4 |
| SEG_5_CL_06 | CP1\|CP2\|DP1 | 1 148 | 71 | 10 | 61 |
| SEG_5_CL_07 | DP1\|CP3 | 530 | 37 | 10 | 27 |
| SEG_5_CL_08 | CP1\|CP2 | 603 | 35 | 10 | 25 |
| SEG_5_CM_00 | No offer | 820 | 57 | 28 | 28 |
| *SEG_5_sum* | | *5 779* | *344* | *106* | *239* |

According to the results above the biggest net profit comes from Segment no 4. The second place belongs to Segment no 2, but it is worth reminding that customers in this cluster are natural buyers, thus they are interested in bank products, they purchase them and generate the volume and then finally the income even if they have not received any offer from the bank. On the third, yet medal place, finds itself Segment no 1. The last two

Segments – Segment no 3 and Segment no 5 give some profits, but they are so small, that the best option is going to not to offer these two clusters.

# Bibliography

Anderson, E.W., Fornell, C., Lehmann, D.R., (1994), "Customer satisfaction, market share, and profitability: findings from Sweden", Journal of Marketing, 58 (3), pp. 53-66.

Angrisani, C., (December 2008), "Coffee Aisle Revamp Lifts Sales". Supermarket News.

Apte, C.V., Hong, S.J., Natarajan, R., Pednault, E.P.D., Tipu, F.A. and Weiss, S.M., (2003), "Data-intensive analytics for predictive modelling", IBM Journal of Research & Development, 47 (1), pp. 17-23.

Barlett, J. E., II, Kotrlik, J. W., Higgins, C. (2001), "Organizational research: Determining appropriate samples for survey research". Information Technology, Learning and Performance Journal, 19 (1) pp.43-50.

Barnett, V., (1991, 2002), "Sample Survey. Principles an methods".

Barney, J., (1991), "Firm resources and sustained competitive advantage", Journal of Management, 17 (1), pp. 99-120.

Belz, C., (1999), "Verkuskompetenz: Chancen in umkampften Markten", (2 ed.), St. Gallen, Wiem: Ueberreuter/Thexis.

Berger, P. and Magliozzi, T. (1992), "The effect of sample size and proportion of buyers in the sample on the performance of list segmentation equations generated by regression analysis", Journal of Direct Marketing, 6 (1), pp. 13-22.

Berger, P., Magliozzi, T., (1992), "The effect of sample size and proportion of buyers in the sample on the performance of list segmentation equations generated by regression analysis", Journal of Direct Marketing, 6 (1), pp. 13-22.

Berger, P.D. and Nasr N.I. (2001), "All theocation of promotion budget to maximize customer equity", Omega 29 (1).

Berger, P.D. and Nasr-Bechwati, N, (2001), " All theocation of promotion budget to maximize customer equity", Omega, 29 (1), pp. 49-61.

Berger, P.D., Bolton, R.N., Boeman, D., Briggs, E., Kumar, V., Parasuraman, A. and Terry, C., (2002), "Marketing actions and the value of customer assets. A framework for customer asset management", Journal of Service Research, 5 (1), pp. 39-54.

Berry, M., Gordon, L., (2000), "Mastering Data Mining. The Art and Science of Customer Relationship Management", John Wiley & Sons, Inc, New York.

Berry, Michale J.A., Gordon S. Linoff (2004), "Data Mining Techniques For Marketing, Sales and Customer Relationship Management". Wiley Computer.

Berson, A., Smith, S., Thearling, K., (2000), "Building Data Mining Applications for CRM", McGraw Hill, New York 2000.

Blattberg, R., Getz, G., Thomas, J.S. (2001), "Customer equity: Building and managing relationships as valuable assets", Harvar Business School Press, Boston MA.

Blattberg, R., Getz, G., Thomas, J.S., (2001), "Customer equity: Building and managing relationships as valuable assets", Harvard Business School Press, Boston MA.

Bohling, T., Bowman, D., LaValle, S., Mittal, V., Narayandas, D., Ramani, G., Varadarajan, R., (2006), "CRM implementation: effectiveness issues and insights", Journal of Service Research, 9 (2), pp. 184-194.

Brachman, R., Anand, T., (1996), "The process of knowledge discovery in databases: a human-centered approach", Menlo Park, CA: AAAI Press, pp.37-58.

Brooks, F., (1995), "The Mythical Man0Minths", anniversary edn. Reading, MA: Addison-Wesley, pp. 153-160.

Chandon, P., Morwitz, V.G, Reinartz, W.J, (2005), "Do intentions really predict beahaviour? Self-generated validity effects in survey research", Journal of Marketing, 69 (2), pp. 1-14.

Chandon, P., Morwitz, V.G., Reinartz, W.J. (2005), "Do intentions really predict behavior? Self-generated validity effects in Survey Research", Journal of Marketing 69 (2), pp. 1-14.

Chang, H.H., Tsay, S.F. (2004), "Integrating of SOM and K-mean in data mining clustering: An empirical study of profitability evaluation", Journal of Information Management, 11 (4), pp. 161-203.

Chang, H.H., Tsay, S.F., (2004), "Integrating of SOM and K-mean in daa mining clustering: an empirical study of CRM and profitability evaluation", Journal of Information Management, 11 (4), pp. 161-203.

Chapman, P., Clinton J., Kerber, R., Khabaza, T., Reinart, T., Shearer, C., Wirth, R., (2000), "CRISP-DM Step-by-Step Data Mining Guide".

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R., (2000), "CRISPDM 1.0 step-by-step data mining guide", Technical report, CRISP-DM.

Cios, K., and Kurgan, L., (2005), "Trends in data mining and knowledge discovery", in Pal, N., and Jain, L. (eds), "Advanced Techniques in Knowledge Discover and Data Mining", Springer, pp. 1-26.

Collis, D.J., Montgmery, C.A., (1995), "Competing on resources: Strategy in the 1990s", Harvard Business School, 73 (4), pp. 118-128.

Coviello, N.E, Brodie, R.J., Danaher, P.J., Johnston, W.J., (2002), "How firms relate to their markets: an empirical examination of comtemporary marketing practices", Journal of Marketing, 66 (3), pp. 33-46.

Davenport, T., H., (February 2009), "How to Design Smart Business Experiments"., Harvard Business Review.

David, F.N. (1949), "Probability Theory for Statistical Methods." Cambridge University Press.

Davies, A., Brady, T., Hobday, M., (2007), "Organizing for Solutions: Systems Seller vs. System Integrator", Industrai Marketing Management, 36 (2), pp. 183-193.

Day, G.,S., (2006), "Aligning the organization with the market", MIT Sloan Management Review, 48 (1), pp. 41-49.

Deighton, J. et al., (1996) "The Future of Interactive Marketing", Harvard Business Review, vol. 11-12.

Devore, J., L., (2008), "Probability and Statistics for Engineering and Scieces, Edition 7", Chapter 9.

Dokoohaki N.,Matskin, M. Personalizing Human Interaction through Hybrid Ontological Profiling: Cultural Heritage Case

Dre, J.H., Mani D.R., Betz, A.L. and Datta, P. (2001), "Targeting customers with statistical and data-mining techniques", Journal of Service Research, 3 (3), pp. 205-219.

Drew, J.H., Mani, D.R., Betz, A.I., Datta, P., (2001), "Targeting customers with statistical and data-mining techniques", Journal of Service Research, 3 (3), pp. 205219.

Drozdenko, R.G., Drake, P.D. (2002), "Optimal database marketing: Strategy, development and data mining", Sage, London.

Drozdenko, R.G., Drake, P.D., (2002), "Optimal database marketing. Strategy development, and data mining. Sage, London.

Duclos, P., Luzardo, R., Mirza, Y., H., (2008), "Refocusing the sales force for cross-sell", McKinsey Quarterly (1), pp.13-15.

Dwyer, P.R., Schorr, P.H., Sejo, O., (1987), "Developing buyer-seller relationships", Journal of Marketing, 51 (2), pp. 11-27.

Dyche, J. and Tech, J., (2001), "The CRM Handbook: A business Guide to Customer Relationship Management", Addison-Wesley, Reading, MA.

Fader, P.S., Hardie, B.G.S., (2009), "Probability model for customer-base analysis", Journal of Interactive Marketing, 23 (1), pp. 61-69.

Fader, P.S., Hardie, B.G.S., Lee, K.L., (2006), "More than meets the eye", Marketing Research, 18, pp. 9-14.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., (1996), "Knowledge discovery and data mining: towards a unifying framework", in Proceedings of the 2nd International Conferene in Knowledge Discovery and Data Mning, Portland OR, pp. 82-88.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996), "From data-mining to knowledge discovery: an overview", in Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., (eds), "Advances in Knowledge Discovert and Data Mining", AAAI Press, pp. 1-34.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., (1996), "From data mining to knowledge discovery in databases", AI Magazine 17 (3), pp. 37-54.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., (1996), "The KDD process for extraction useful knowledge from volumes of data", Communications of the ACM, 39 (11), pp. 27-34.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., (1996), Handbook "Advances in Knowledge Discovery and Data Mining",AAAI Press, pp. 1-34.

Felvey, J., (1982), "Cross-selling by computer", Bank Marketing, (July), pp. 25-27.

Fishman, Ch., (2009), "This is a Marketing Revolution". Fast Company.

Fleenor, D. Gail (June 2009), "These Tests Are Positive". STORES Magazine.

Fleming, C., (2006), "Secrets to cross-selling success", Credit Union Magazine Vol. 72: Credit Union National Association, Inc.

Foote, N., W., Galbraith, J., Hope, Q., Miller, D., (2001), "Making solutions the answer", McKinsey Quarterly (3), pp. 84-93.

Frawley, W., Piatetsky-Shapiro, G., and Matheus, C., (1991), "Knowledge discovery in databases", AAAI/MIT Press, pp. 1-27.

Freund, J., E. (1984) „Modern Elementary Statistics". Handbook, Prentice hall (Chapter "Significance Tests").

Galbraith, J.R., (2005), "Designing the customer-centric organization: a guide to strategy, structure and process", Handbook by A Wiley Imprint.

Ganesh, J., Arnold, M.J., Reynolds, K.E., (2000), "Understanding the customer base of service providers: An examination of the differences between switchers and stayers", Journal of Marketing, 64 (3), pp. 65-87.

Glazer, R., (1997), "Strategy and structure in information - intensive markets: the relationship between marketing and IT", Journal of Market – Focused Management, 2 (1), pp. 65-81.

Grant, R.M., (1991), "The resource-based theory of competitive advantage: implications for strategy formulation", California Management Review, 33 (3), pp. 114-135.

Gulati R., Oldroyd, J., B., (2005), "The quest for customer focus", Harvard Business Review, 83 (4), pp. 92-101.

Gulati, R., (2007), "Silo busting", Harvard Business Review, 85 (5), pp. 98-108.

Gupta, S., Lehmann, D.R., (2003), "Customers as assets", Journal of Interactive Marketing, 17 (1), pp. 9-24.

Gupta, S., Zeithaml, V., (2006), "Customer metrics and their impact on financial performance", Marketing Science, 25 (6), pp. 718-739.

Gurau, C., Ranchhod, A., (2002), "How to calculate the value of a customer", Journal of Targeting, Measurement and Analysis for Marketing, 10 (3), pp. 20-220.

Han, J., Kamber, M., (2006), "Data Mining. Concepts and Techniques". Handbook by Elsevier.

Harding, F., (2004), "Cross selling or Cross Purposes?", Harvard Business Review, July-August 2004, pp. 45-49.

Hobby, J., (1999), "Looking After the One Who Matters," Accountancy Age, (October 28), 28–30.

Homburg, C., Kuester, S., (2001), "Towards as improved understanding of industrial buying behavior: determinants of the number of suppliers", Journal of Business-to-Business Marketing, 8 (2), pp. 5-32.

Hosmer, D.W., Lemeshow, S., (2000), "Applied logistic Regression".

Hu, W., Jing, Z., (2008), "Study of customer segmentation for auto services companies based on RFM model", Proceedings for ICIM 2008: International Conference on Innovation and Management.

Jackson, D., (1989), "Determining a customer`s lifetime value (Part2)", Direct Marketing, 52 (1), pp. 24-32.

Jackson, D., (1989), "Determining a customer`s lifetime value", Direct Marketing, 51 (11), pp. 60-62.

Jackson, D., (1989), "Insurance marketing: Determining a customer`s lifetime value (Part3)", Direct Marketing, 52 (4), pp. 28-30.

Jain, D., Singh, S., (2002), "Customer Lifetime Value Research in Marketing: A Review and Future Directions", Journal of Interactive Marketing, 16 (2), pp. 34-46.

Kamakura, W., Ramaswani, S., Srivasta, R., (1991), „Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services", International Journal of Research in Marketing, 8 (4), pp. 329-349.

Kamakura, W., Wedel, B., Rosa F.D., Mazzon, J.A., (2003), „Cross-selling through database marketing: A mixed factor analyzer for data augmentation and prediction", International Journal of Research in Marketing, 20 (1), pp. 45-65.

Kantardzic, M., (2011), "Data Mining. Concepts, Models, Methods and Algorithms." Handbook by John Wiley & Sons.

Khanna, S., (2001), "Measuring the CRM ROI: show them benefits". Retrieved from http://www.crm-forum.com.

King, H., Honaker, J., Joseph, A., and Scheve, K., (2001), "Analyzing incomplete policitval science data: An alternative algorithm for multiple imputation", American Political Science Review, 95 (1), pp. 49-69.

Kish, L (1965), "Survey Sampling". Wiley Handbook.

Klosgen, W., (1992), "Problems for knowledge discovery in databases and their treatment in the statistics interpreter explora", Kournal of Intelligent Systems, 7 (7), pp. 649-673.

Klosgen, W., and Zytkow, J., (1996), "Knowledge discovery in databases terminology", in Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., (eds), "Advances in Knowledge Discovery and Data Mining", AAAI Press, pp. 573-592.

Knott, A., Hayes, A., Neslin, S., (2002), "Next-product-to-buy models for cross-selling applications", Journal of Interactive Marketing, 16 (3), pp. 59-75.

Kohli, A.K., Jaworski, B.J., (1990), „Market orientation: the construct, research, proposition and managerial implications", Journal of Marketing, 54 (2), pp. 1-18.

Kumar V.,(2010) „Zarządzanie Wartością Klienta", WP PWN, Warszawa.

Kumar, V., George, M., Pancras, J., (2007), "Cross-buying in retailing: Drivers and consequences", Journal of Retailing, 84 (1), pp. 15-27.

Kumar, V., Rajan, B., (2009), "Nurturing the right customers", Strategic Finance, September, pp. 27-33.

Kumar, V., Ramani, G., Bohling, T., (2004), „Customer lifetime value approaches and best practice applications", Journal of Interactive Marketing, 18 (3), pp. 60-72.

Kumar, V., Shah, D., (2004), "Building and sustaining profitable customer loyalty for the 21st century", Journal of Retailing, 80 (4), pp. 317-330.

Kumar, V., Shah, D., (2009), "Expanding the role of marketing: from customer equity to marketing capitalization", Journal of Marketing, 73 (6), pp. 119-136.

Kumar, V., Venkatesan, R., Reinartz, W., (2008), "Performance implications of adopting a customer-focused sales campaign", Journal of Marketing, 72 (5), pp. 50-68.

Kurgan, L., Musilek, P., (2006), "A survey of Knowledge Discovery and Data Mining process models", The Knowledge Engineering Review, 21 (1), pp. 1-24.

LaPlaca, P.J., (2004), "Letter from the editor: Special issue on customer relationship management", Industrial Marketing Management, 33 (6), pp. 463-464.

Larose, D., T., (2005), "Discovering Knowledge in Data: An Introduction to Data Mining". Handbook by John Wiley & Sons.

Larose, D.T., (2005), "Discovering knowledge in data: an introduction to data mining", Handbook by Wiley and Sons Inc.

Lazer D. et al., (2009) " Life in the network: the coming age of computational social science", Science, vol. 6, 323 (5915), 721-723.

Levitt, T., (1983), "After the Sale is Over….", Harvard Business Review, 61 (September/October), pp. 87-93.

Levitt, T., (1983), "After the sale is over…", Harvard Business Review, 61, pp. 87-93.

Maimon, O., Rokach, L., (2005), "Data Mining and Knowledge Discovery Handbook", by Springer Inc.

Maimon, O., Rokach, L., (2005), "The Data Mining and Knowledge Discovery Handbook". Handbook by Springer Science and Business Media, Inc.

Marbán, Ó., Mariscal, G., Segovia, J., (2009), "A Data Mining & Knowledge Discovery process model", in Ponce, J., and Karahoca A., "Data Mining and Knowledge Discovery in real life applications", I-Tech, Austria.

Marbán, Ó., Segovia, J., Menasalvas, E. and Fernandez-Baizan, C., (2008), "Towards Data Mining Engineering: a software engineering approach", Information systems Journal.

McCarty, J.A, Hastak, M., (2007), "Segmentation approaches in data-mining: A comparison of RFM, CHAID and logistic regression", Journal of Business Research, 60 (6), pp. 656-662.

Menard, S., (2001), "Applied Logistic Regression Analysis".

Mitussis, D., O`Malley, L., and Patterson, M., (2006), "Mapping the Re-engagement of CRM with Relationship Marketing", European Journal of Marketing, 40 (5/6), pp. 572-589.

Morgan, R.M., Hunt, S.D., (1994), "The commitment-trust theory if relationship marketing", Journal of Marketing, 58 (3), pp. 20-38.

Mundt, K., Dawes, J., Sharp, B., (2006), "Can a brand outperform competitors on cross-category loyalty? An examination of cross-selling metrics in two financial services markets", Journal of Consumer Marketing, 23 (7), pp. 465-469.

Ngai, E.W., Xiu, L., Chau, D.C.K., (2009), "Application of data mining techniques in customer relationship management: A literature review and classification", Expert systems with Applications, 36 (2), pp. 2292-2602.

Palmatier, R., W., (2008), "Relationship Marketing", Handbook of Marketing Science Insitute, Relevant Knowledge Series.

Palmatier, R.W., Sheer, L.K., Evans, K.R., Arnold, T.J., (2008), "Achieving relationship marketing effectiveness in business-to-business exchanges", Journal of Marketing Research, 44 (2), pp. 185-199.

Parvatiyar, A., and Sheth, J.N., (2001), "Customer relationship management: Emerging practice, process and discipline", Journal of Economic and Social Reseach, 3 (2), pp. 1-34.

Paulin, M., Perrien, J., Ferguson, R.J., Salazar, A.M.A., Seruya, L.M., (1998), "Relation norms and client retention: External effectiveness of commercial banking in Canada and Mexico", International Journal of Marketing, 69 (4), pp. 167-176.

Payne, A., and Frow, P., (2005), "A strategic framework for customer relationship management", Journal of Marketing, 69 (4), pp. 167-176.

Pfeifer, P., and Caraway, R., (2000), "Modeling customer relationship as Markov chains", Journal of Interactive Marketing, 14 (2), pp. 43-55.

Piatetsky-Shapiro, G., (1991), "Knowledge discovery in real databases: a report on the IJCAI-89 workshop", AI Magazine, 11 (5), pp.. 68-70.

Piatetsky-Shapiro, G., and Matheus, C., (1992), "Knowledge discovery workbench for exploring busness databases", International Journal of Intelligent Agents, 7 (7), pp. 675-686.

Piller, F., (2011), "The market fir customization and personalization – where will we go?", CYO Conference – create your own 2011, Berlin.

Pressman, R., (2005), "Software Engineering: a practitioner`s approach", McGraw-Hill, New York.

Reichheld, F.F., and Sasser, W.E., (1990), "Zero defecions: Quality comes to services", Harvard Business Review, 68 (5), pp. 105-111.

Reichheld, F.F., and Teal, T., (1996), "The loyalty effect: The hidden force behind growth, profits, and lasting value" Harvard Business School Press, Boston, Massachusetts.

Reinartz, J.S., Thomas, J.S., and Bascoul, G., (2008), "Investigating cross-buying and customer loyalty", Journal of Interactive Marketing, 22 (1), pp. 5-20.

Reinartz, T., (2002), "Stages of the discovery process", from Klosgen, W., Zytkow, J., "Handbook of Data Mining and Knowledge Discovery", Oxford Univeristy Press, pp. 185-192.

Reinartz, W. and Venkatesan, R., (2008), "Decision models for customer relationship management (CRM)", in Wierenga, B., (2008). Handbook of Marketing Decision Models, International Series in Operations Research and Management Science, Volume 121, Part IV, pp. 291-326, Springer, Rotterdam, US.

Reinartz, W., Thomas, J.S., Bascoul, G. (2008), "Investigation cross-buying and customer loyalty", Journal of Interactive Marketing, 22 (1), pp. 5-20.

Reinartz, W., Venkatesan, R. (2008), "Decsion models for Customer Relationship Management (CRM)", in Wiereng, B. (2008), Handbook of Marketing Decision Models, International Series in Operations Research & Management Science, Volume 121, Part IV, pp. 291-326, Springer, Rotterdam, US.

Richard, K.A., Jones, E., (2008), "Customer relationshio management: fnding value druvers", Industri Marketing Management, 37 (2), pp. 120-130.

Rigby, D.K., Reichheld, F.F., Shefter, P., (2002), "Avoid the four perils of CRM", Harvard Business Review, 80 (2), pp. 101-109.

Rouse, W., (2002), "Need to know-information, knowledge, and decision making", IEEE Transactions on Systems, Man and Cybernetics, Part C, 32 (4), pp. 282-292.

Rubin, D.B., (1976), "Inference and missing data", Biometrica, pp. 581-592.

Rust, R.T., Chung, T.S. (2006), "Marketing models of service and relationships", Marketing Science, 25 (6), pp. 560-764.

Rust, R.T., Chung, T.S., (2006), "Marketing models of service and relationships", Marketing Science, 25 (6), pp. 560-764.

SAS Intitute Inc., (1998), "SAS Technical Report A-108. Clustering Cubic Criterion", Cary, NC, USA.

SAS, 1997, SAS Institute Inc, "From Data to Business Advantage: Data Mining, SEMMA Methodology and the SAS System" (White Paper).

Schäfer, H., (2002), "The development of customer potential through cross-selling", Wiesbaden: Deutscher Universitäts-Verlag / GWV Fachverlage GmbH.

Selden, L, Colvin, G., (2003), "Angel Customers and Demon Customers: Discover Which is Which and, Turob-Charge Your Stock", Handbook by Portfolio.

Seybold, P.B., (2001), "Get inside the lives of your customer", Harvard Business Review, May, pp. 80-91.

Shah, D., Rust, R.T., Parasuraman, A., Staelin, R., Day, G.S., (2006), "The path to customer centricity", Journal of Service Research, 9 (2), pp. 113-124.

Singh, D., Agrawal, D.P., (2003), "CRM Practices in Indian Industries", Internation Journal of Customer Relationship Managemenr, 5, pp. 241-257.

Smith, A., (2006), "CRM and customer service: strategic asset or corporate overhead?", Handbook of Business Strategy, Vol. 7 Iss: 1, pp. 87 – 93.

Srivastava, R.K., Shervani, T.A., Fahey, L., (1998), "Market-based assets and shareholder value: a framework for analysis", Journal of Marketing, 62 (1), pp. 2-18.

Stone, M., Woodcock, N. (2001), "Defining CRM and assessing its quality", in "Successful Customer Relationship Marketing" by Foss, B., Stone, M., (eds), London: Kogan Page, pp. 3-20.

Surma, J., (2012), "Modeling Customer Behavior with Analytical Profiles", Social Network Mining, Analysis and Research Trends: Techniques and Applications, pp. 171-182.

Swift, R., (2000), “Accelerating Customer Relationship – Using CRM and Relationship Technologies”, Upper Saddle River, NJ: Prentive Hall.

Swift, R.S., (2001), “Accelerating customer relationship: using CRM and relationship technologies”, Prentice Hall, Upper Saddle River, NJ.

Tedlow, R. S., Jones, G., (1993), “The rise and fall of mass marketing”. Handbook.

Tuli, K.R., Kohli, A.K., Bharadwaj, S.G, (2007), „Rethinking customer solutions: from product bundles to relational processes”, Journal of Marketing, 71 (3), pp. 1-17.

Van den Poel, D., Lariviere, B., (2004), “Customer attrition analysis for financial services using proportional hazard models”, European Journal of Operational Research, 157 (1), pp. 196-217.

Wernerfelt, B., (1984), “A resource-based view of the firm”, Strategic Management Journal, 5 (2), pp. 171-180.

Wiederhold, G., (1996), “Foreword: on the barriers and future of knowledge discovery”, in Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., ad Uthurusamy, R., (eds), “Advances in Knowledge Discovery and Data Mining”, AAAI Press.

Winer, R.S., (2001), “A framework for customer relationship management”, California Management Review, 43 (4), pp. 89-105.

Witten, I., H., Frank, E., (2005), “Data Mining: Practical Machine Learning Tools and Techniques”. Handbook by Elsevier.

Witten, I.H., Frank, E., (2005), “Data Mining – Practical Machine Learning Tools and Techniques”, Handbook by Elsavier Inc.

Wong, E., (June 2009), "Case Study: How Search Ads Helped Pier 1 Stay Afloat". Brandweek.

Zablah, A.R., Beuenger, D.N., Wesley J.J, (2003), “Customer Relationship Management: an explication of its domain and avenues for further inquiry”, in “Relationship Marketing, Customer Relationship Managemen nad Marketing Managemenr: Co-Operation-Competition-Coevolution”, Kleinaltenkamp M., Ehret, M., eds. Berlin: Freie Universitat Berlin, pp. 115-124.

Zairko, R., Golan, R., and Edwards, D., (1993), “An application of Datalogic/R knowledge discovery tool to identify strong predictive rules in stock market data”, working notes from the Workshop on Knowledge Discovery in Databases, Seattle, Washington, pp. 89-101.

Zytow, J., and Baker, J., (1991), “Interactive mining of regularities in databases”, in Piatetsky-Shapiro, G., and Frowley, W.J., (eds), “Knowledge Discovery in Databases”, AAAI Press, pp. 31-53.