ZESZYT NAUKOWY
Wyższa Szkoła Zarządzania i Bankowości w Krakowie

Jan J. Juras
Dr Marek A. Valenta
The School of Banking and Management in Krakow
juras.jan@gmail.com, valenta@agh.edu.pl

OPTIMIZATION OF ONLINE MARKETING – SALES FORECASTING WITH THE APPLICATION OF LINNEAR REGRESSION

Introduction

The development of IT and the omnipresence of the Internet in everyday life activities result in the necessity to adapt the services offered to wide numbers of customers to these technologies. In order to achieve a competitive advantage among service providers, new different methods of advertising must be applied in that new environment[1]. Thanks to the existence of on-line advertising the possibility to react promptly to the changes in value of particular process parameters that have a significant impact on the achievement of business targets is their new and most crucial feature. Thus, the following tasks became significant: the differentiation of important parameters that describe business processes, the determination of the character of their impact on the achievement of business targets and the development of an IT system to support decision-making by the managers of the process. The paper presents the results of the research[2] whose aim was to achieve the above listed three tasks dedicated to a selected business process that was typical for services related to a wide range of physical recreation.

1. Problem formulation and the concept of the on-line marketing optimization support

[1] Hennig-Thurau T. et al., *The Impact of New Media on Customer Relationships*, Journal of Service Research, vol. 13, (pp. 311-330), 2010.
[2] Juras J., *Optymalizacja reklamy w Internecie z wykorzystaniem regresji liniowej,* WSZiB, Kraków 2015 (diploma thesis supervised by Valenta M.)

The problem concerns a particular case of service provision that is supported – which is most important here – by the possibility to run marketing operations in the form of on-line advertising.

The process under analysis concerned the acquisition of course participants by a service company that provided a variety of yoga courses. The participants were acquired by a wide range of information activities which consisted mainly in online marketing. Apart from the fact that online marketing practices can adopt various forms, they definitely differ from other types of advertising as their intensity can be easily controlled by changing the form and frequency of appearance on the web[3]. It is important that in the key moment of the enrollment campaign the operation could be effectively influenced on the basis of the knowledge of the current effects. The control of the intensity of the marketing campaign on the basis of its current results is not possible in the case of other than online forms of marketing as they require much earlier planning operations.

The case under investigation concerns the enrollment of yoga class participants but in fact it involves the whole range of leisure and recreational services. Thus, the solution of the presented problem will provide solutions that are useful in the wide range of services similar in character[4].

The knowledge obtained from managers who had experience in the sector suggested definitely that the final level of the sales of the service (the level of the business target achievement) depends on the total sales of the service on the days directly preceding the beginning of the course or even – due to the character of the service offered – on the course commencement date or on the days after[5].

The optimization of marketing operations should be supported by decisions that are either made by experienced managers or by a decision support system (preferably electronic) that is dedicated to such tasks. The dedication of the decision–support system consists in its precise adaption to the character of business activity that is frequently difficult to define; not speaking of measuring and controlling its parameters. Consequently, it is almost impossible to develop an algorithm of a decision-making system on the basis of detailed knowledge with numerous indefinitely defined parameters.

---

[3] Dye R., *The Buzz on Buzz,* "Harvard Business Review" 2007, www://hbswk.hbs.edu/archive/1956.html.
[4] Wiktor, J.W., *Promocja usług*, [in:] Czubała A., Jonas, A., Smoleń T., Wiktor J.W. *Marketing usług* (pp. 213-255), Oficyna Ekonomiczna, Kraków 2006.
[5] J. Juras, *Optymalizacja…,* op. cit.

The presented here idea of developing a decision-making support system as regards the control of the degree of marketing campaign intensity assumes the possibility to construct a system that would take advantage of the heuristic knowledge gained from learning by examples.

In the presented case, the system is based on the possibility to apply a significant amount of data concerning marketing campaigns and the business targets reached that was obtained during numerous courses advertised in recent years. The available amount of data is representative and it can be used to define the behavior rules of the reality that we are going to control in the future through decisions that concern the intensity of on-line advertising.

The possibility to define such rules constitutes the condition for developing a computer system that makes it possible to conclude on the final achievement of business targets on the basis of the knowledge on the level of the current sales of the service. That may constitute the basis for current decisions regarding the control over the intensity of an on-line marketing campaign. Chapter 2 presents a method of acquiring the fundamentals for the development of the reality behavior rules, while chapter 3 presents a design and implementation remarks concerning the development of a computer system (a prototype) that would enable the user an ongoing collection of data on business targets achieved and to predict the final effect of the campaign at the end of the enrollment campaign. As it was mentioned before, such forecasts facilitate a flexible adaptation of the form and intensity of advertising operations in the last (and the most crucial) days of the campaign, which optimizes the costs involved and increases the probability of achieving the defined business targets. By *the last days* of the advertising campaign the authors mean two days before and two days after the commencement of the course (5 days in total).

2. Developing business rules with the application of linear regression method

In order to optimize the control of processes in business environment, it is frequently necessary to forecast the values of parameters that are significant (e.g. the demand for the service, the sales volume or the costs). The determination of crucial parameters that describe company functioning such as e.g. the average sales volume or maximum and minimum sales volume is usually not difficult for company managers. However, it is more difficult to define the character and parameters of relationships between various process parameters that practically can be determined with the application of more or less complicated analytical methods that require, for example, statistical computations. They may be helpful in determining

the impact of particular external and internal factors on company financial results and in many cases they are useful in the determination of – for example – the optimum advertising expenses or in the assessment of the impact of the price and the volume of advertising investment on sales[6].

Pearson correlation coefficient method and the least square linear regression analysis were applied in the investigations in order to obtain statistically significant relationships in the selected service sales process so that business rules could be formulated for the implementation of the optimization of advertising intensity in order to achieve the defined sales target[7].

One of the Krakow yoga schools served as the example of the research environment. Thanks to many years of functioning, the school stored and provided the researchers with a set of sales data on numerous enrollment campaigns that were run in recent years.

The first important step of the research was to select variables whose values would be available in the enrollment process and which according to enrollment managers might determine the final level of successful enrollment (the level of the defined business target). Moreover, it was concluded from the experience of one of the authors[8] that prompt reaction and adequate changes in the intensity of marketing campaigns with regard to the change in sales a few days before the start of the course and during the first two – five days of the course were crucial to the final sales/recruitment results. These are the days on which on-line advertising campaigns should dynamically and efficiently influence the final recruitment result.

Thus, graphs were developed to present the distribution of the variables under investigation, and the character of the phenomena observed led to the conclusion that the correlation of the variables was linear. The acceptance of the linear regression as the description method of the expected correlations resulted finally in obtaining comparatively simple and easy prediction rules to be implemented in the future system. It was assumed that the explanatory variables are the volumes of sales on the course commencement day, on two and one days before and two and one days after the beginning of the course (x_3, x_1 and x_2, and x_4 and x_5, respectively), while the final sales of the services was the dependent variable y. Through the least square linear regression analysis for the isolated pairs of variables, linear regression coefficients b and a (y=bx+4) were obtained. Another step was to determine the significance

[6] Kotler P., Armstrong G., Saunders J., Wong V., *Marketing Podręcznik Europejski*. Polskie Wydawnictwo Ekonomiczne, Warszawa 2002.
[7] Nowak E., Analiza statystyczna w ocenie działalności przedsiębiorstwa [in:] Nowak E. (ed.), Metody statystyczne w analizie działalności przedsiębiorstwa, Polskie Wydawnictwo Ekonomiczne, Warszawa 2001.
[8] Juras J. *Optymalizacja…*, op. cit.

level of dependencies between the accepted variables. Here, the Pearson coefficient correlation method was applied.

A list of 5 independent variables (denoted by x_1 - x_5) and the dependent variable y were selected for the investigation:

x_1 - sales result obtained 2 days before the course commencement day,

x_2 - sales result obtained 1 day before the course commencement day,

x_3 – sales result obtained on the course commencement day,

x_4 – sales result obtained one day after the course commencement day,

x_5 – sales result obtained two days after the course commencement day.

Y – the sales result achieved at the end of the enrollment process that finishes one week at the latest after the service started.

It was assumed that the monitoring of variables x_i should make it possible – with the help of a decision support system - to take optimal solutions as regards the control of the intensity of on-line marketing operations on the course commencement day, two and one days before and two and one days after it.

Further research aimed at confirming two hypotheses (for each of variables x_i):

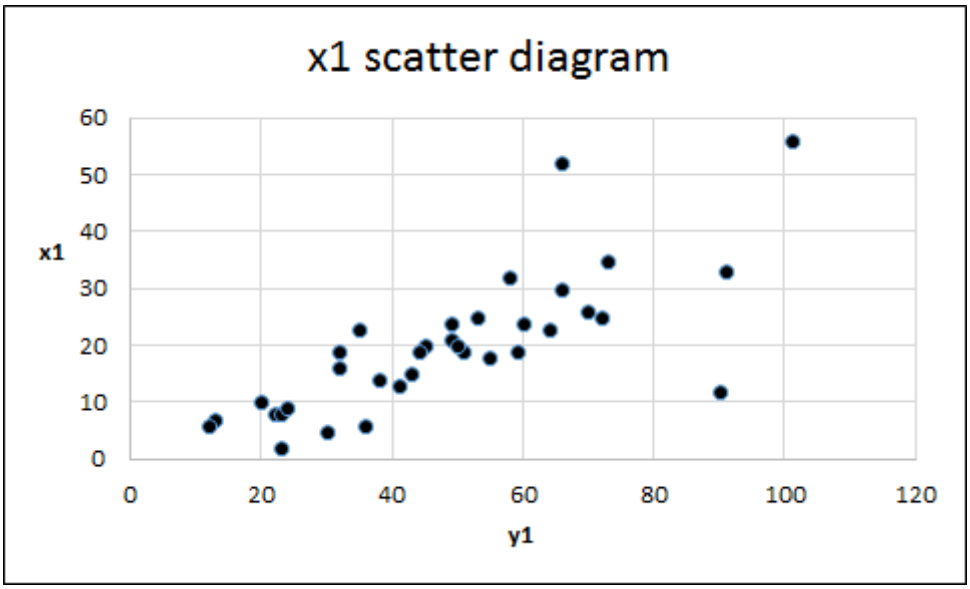
- sales result x_i achieved on a particular day of sales before/after the course commencement is significantly correlated with the final sales result y,
- it is possible to predict final sales result y on the basis of the sales result achieved on a particular day of sales before/after the course commencement.

The verification of the above hypotheses was conducted on the basis of factual sales data from previous years. For each of variables x_i Pearson linear correlation coefficient r was determined for pairs x_i (the sales on a particular day) and y (the final sales of the service); the parameters of appropriate models in the form of coefficients a and b were defined with the application of the linear regression analysis (the least square method); and the standard error was estimated.

The scatter diagram for variable x_1 (the sales obtained two days before the service started) and the values for final sales y are given in figure 1. The scatter diagrams for subsequent variables x_2–x_5 and y were similar.

Table 1 presents the results of the correlation analysis for variables x_2–x_5 and y, the values of Pearson correlation coefficients r and p-values from the statistical test, respectively.

Figure 1. Scatter diagram for variable x_1



Source: Juras J., *Optymalizacja reklamy w Internecie z wykorzystaniem regresji liniowej*, WSZiB, Kraków 2015 (diploma thesis, supervised by Valenta M.)

Table 1. Result of correlation analysis for variables $x_1 – x_5$ and y

	r (Pearson)	p-value
x_1	0,7655	0,00000008
x_2	0,7922	0,0000000014
x_3	0,8229	0,0000000013
x_4	0,8777	0,15
x_5	0,7922	0,19

Source: Juras J., *Optymalizacja reklamy w Internecie z wykorzystaniem regresji liniowej*, WSZiB, Kraków 2015 (diploma thesis, supervised by Valenta M.)

It can be concluded from the data obtained that there are "moderately strong" correlations between x_1 and y and x_2 and y that are statistically significant ($p < 0,05$).

In the case of x_3 and y, x_4 and y, and x_5 and y there are "strong correlations" that are also statistically significant ($p < 0,05$).

After computations, on the basis of the above mentioned historical data, the values were obtained of linear regression parameters a and b that model the dependencies of y from x_1-x_5. Moreover, for each of the variables the values of standard error and statistical significance p were determined. The results of the computations are given in table 2 and the tools applied are presented in chapter 3 as an element of the system implementation.

Table 2. Results of the linear regression analysis of variables x_1-x_5 and y

	factor b	factor a	SD	p-value
x_1	1,41	20,36	4,75	0,0001
x_2	1,33	15,19	4,92	0,0002
x_3	1,13	12,14	4,85	0,017
x_4	1,16	6,41	4,37	0,15
x_5	1,31	5,66	4,22	0,19

Source: Juras J., *Optymalizacja reklamy w Internecie z wykorzystaniem regresji liniowej,* WSZiB, Kraków 2015 (diploma thesis, supervised by Valenta M.)

Regression models for the pairs of variables x_1 and y, x_2 and y, and x_3 and y are statistically significant ($p < 0,05$), while for pairs x_4 and y, x_5 and y, the models obtained are insignificant ($p > 0,05$).

For variables x_1, x_2 and x_3 the hypotheses were confirmed that it is possible to predict the final sales results on the basis of the sales result on two and one days before the course commencement and on the day of its commencement, respectively.

In the case of variables x_4 and x_5, the hypothesis that the prediction of the final sales result is possible on the basis of the sales result on one and two days after the commencement of the course were not confirmed due to the lack of the statistical significance of the model obtained.

Finally, after the verification of the hypotheses, 3 business rules were developed according to which the sales results achieved two and one day before and on the day when a service was commenced enable forecasting the final sales volume.

In accordance with the data given in table 2, the equations of regressions for the hypotheses that were confirmed and for the related rules are:

$y = 20,36(+/- 4,75) + 1,41*x_1$

$y = 15,91(+/- 4,92) + 1,33*x_2$

$y = 12,14(+/- 4,85) + 1,13*x_3$

The knowledge on the behavior of reality presented in the form of the above rules that were obtained on the basis of the sales database from recent years constituted the basis for the development of a system prototype whose characteristics is given below.

3. Support system for cost optimization of marketing decisions

The possibility was confirmed to forecast the total final sales of a service on the basis of the sales that were achieved two days before, one day before and on the day of the service commencement. The basic function of the system is to support decision-making processes of the management of a company in the recreational sector that concern the optimization of the advertising and marketing operations that regard courses organized by the company. It is assumed by the system that the management defines the sales target for a particular service in the form of the number of its participants.

In the course of the sales campaign, the database of the system collects the data on the service sales. On the basis of the data collected, with the application of the regression equations that were determined on the basis of all historical data in the database, the forecast for the total sales in the current enrollment campaign is determined and compared to the defined sales target. The forecast given by the system and the information whether the sales target will be achieved constitute the basis for a possible modification of advertising activities. In the case when the predicted sales volume diverges from the target, the management can promptly (as it is done on-line) adjust the intensity of marketing operations to the forecast. On the other hand, if the predicted sales exceed the assumed volume and the forecast indicates that the capacity to supply the service is exhausted, the management may withdraw the advertising from the Internet to avoid the generation of unnecessary costs.

It is assumed that the system should additionally have the capacity to adapt to future reality. The initial calculation of the parameters for the rules was conducted on the basis of historical data. However, the adaptation character of the system can be achieved by:

- an on-going extension of the database with the sales results of subsequent services and marketing operations that accompanied them,
- an on-going calculation– with the application of the system – of the new values of rule parameters on the basis of the up-dated sales database.

The development of the system that was described in more detail in [4] considered one actor/user that operates the system, i.e.:

- adds new data regarding new services (courses),
- adds “book now” function to new services (courses),
- services customers as regards the choice of services and their reservation,

- calculates recommendations for possible modifications of the advertising campaign intensity,
- has access to the analysis results of sales figures and the information about services.

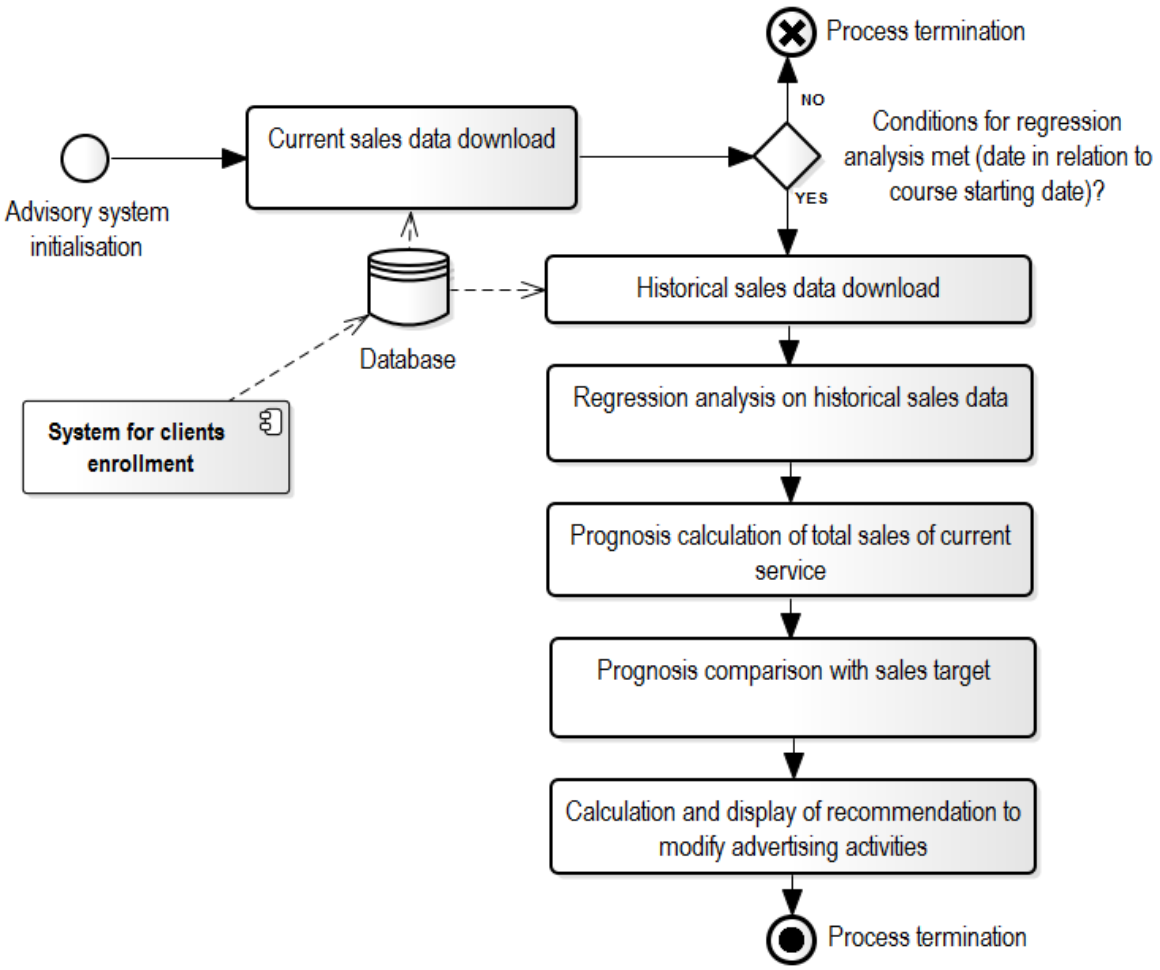
The system was written in Java 1.6, JavaEE 1.6 and run on JBOSS (7.1) application server. The software architecture was MVC-based. JSF technology and PrimeFace 2.1library were applied to implement the presentation level.

When developing the system, the following software was also applied: Spring (relationships download), MySQL (database), Hibernate (ORM), JUnit (tests), Maven, Git (version control); and org.apache.commons.math3.stat.regression[9] library to apply the functions related to the rule development (the coefficients of the least square method of linear regression).

A graphic presentation of the whole processing operation performed by the system is given in the BPMN diagram in fig.2.

[9] Library: org.apache.commons.math3.stat.regression http://commons.apache.org/proper/commons-math/apidocs/org/apache/commons/math3/stat/regression/package-summary.html

Figure 2. BPMN diagram for determining the recommendations to change the intensity of marketing activities



Source: Authors research

The role of the Database in the system cannot be overestimated. It stores the historical information on previous services and the enrollment activities related to them as well as it stores and updates data on the current service and its all bookings.

The most important data concern the name of the service, the date of its provision, the maximum number of participants, a defined sales target and the service sales volume in the particular days of the recruitment and (simultaneously) the advertising campaigns.

The above data, together with the information on the current date are the basis for passing to and performing the subsequent steps of the system operation

The subsequent steps of processing are performed when the current system date equals the date of the service commencement or when it precedes it by one or two days i.e. the days of the recruitment campaign that were indicated by the procedures described in chapter 2.

The most significant functions of the system are as follows:

- download from the database the updated historical data that concern the sales of the services,
- develop business rules after the regression analysis of the downloaded data,
- calculate, on the basis of the rules, the predicted level of the service sales based on the hitherto accumulated sales on the day of the service commencement or one or two days before, respectively.

Further functions of the system concern only its advisory role:

- comparing the calculated sales forecast with the defined sales target,
- presenting the recommendations to modify the intensity of marketing operations.

Conclusion

The article presents the investigation results of a business process regarding the sales of services (the participation in a course) in a company in the physical recreation and leisure sector (yoga courses) whose advertising campaign is conducted mostly on-line. Such form of campaign makes it possible to react quickly as regards the alterations of its intensity, which has an impact on company costs.

The solutions to the following two tasks are given in the article:

- determine the possibility and method how to draw conclusions from current sales results as regards the final result of marketing operations,
- develop a tool that would support decision-making with regard to the cost optimization of online marketing operations in order to achieve the business target (the forecast volume of sales)

It should be admitted that the second task was achieved[10] with respect to the development of the system prototype. The prototype ignored the function that is generally referred to as the "system administration" as it was considered irrelevant. In the authors' opinion one more function that was not implemented in the prototype (and is designed and visible in the BPMN diagram – Fig.2) may decide on the quality and practical usefulness of the developed system. It may be crucial when the system is used in companies that offer services different than the

[10] Juras J., *Optymalizacja…,* op. cit.

ones in the recreational sector (e.g. in tourist and educational sectors which are seasonal in character).

It is typical for such sectors that the demand for their services (as well as the enrollment and the sales) depends on the time of year when they are offered. Generally, the sale of services in this sector is extremely seasonal and in some months the demand increases (e.g. the biggest demand in the recreational sector is in autumn and winter months). Thus, it is important when a course is started or service is offered (the day of the week, month, holiday month, summer or winter season, a bank holiday or the so called "long weekend"). One should not forget about the impact of weather conditions on the day when the advertised service starts. The latter factor is extremely difficult to consider without a substantial development of the database or a close cooperation between the system and weather services.

It is worth mentioning that there is a possibility to extend the system by a module that makes it possible to take into consideration the influence of the time of year on sales. In such cases, the database includes historical data with the dates of the recorded recruitment campaigns. Consequently, the module that downloads historical sales data from the database in order to specify the rule dedicated to a service offered on a particular date may do it selectively - i.e. .adequately to the date and situation as suggested by the information from previous years.

The existing prototype was tested in a real life company that made its historical data available. The prototype proved its usefulness as it confirmed the correctness of the final sales forecasts that were based on the sales on the service commencement day and one and two days before.

Bibliography

[1] Biblioteka org.apache.commons.math3.stat.regression http://commons.apache.org/proper/commons-math/apidocs/org/apache/commons/math3/stat/regression/package-summary.html

[2] Dye R., *The Buzz on Buzz*, "Harvard Business Review" 2007, www://hbswk.hbs.edu/archive/1956.html.

[3] Hennig-Thurau T. Et al., *The Impact of New Media on Customer Relationships*, Journal of Service Research, vol. 13, (pp. 311-330), 2010.

[4] Juras J., *Optymalizacja reklamy w Internecie z wykorzystaniem regresji liniowej*, WSZiB, Kraków 2015 (diploma thesis, supervised by Valenta M.).

[5] Kotler P., Armstrong G., Saunders J., Wong V., *Marketing Podręcznik Europejski*. Polskie Wydawnictwo Ekonomiczne, Warszawa 2002.

[6] Nowak E., Analiza statystyczna w ocenie działalności przedsiębiorstwa [in:] Nowak E. (ed.), Metody statystyczne w analizie działalności przedsiębiorstwa, Polskie Wydawnictwo Ekonomiczne, Warszawa 2001.

[7] Wiktor, J.W., *Promocja usług*, [in:] Czubała A., Jonas, A., Smoleń T., Wiktor J.W. *Marketing usług* (pp. 213-255), Oficyna Ekonomiczna, Kraków 2006.

Abstract

The solution presented in the article is dedicated to controlling economic processes in the sector of services related to physical recreation. Thanks to the support of IT and the Internet, the optimization of marketing in the sector may also concern the dynamics of operations in the very short period that precedes the moment when a particular service is rendered and which is often crucial as regards the acquisition of an optimal number of customers. Due to the application of the linear regression method in the investigation of the character of relationships between selected processes, rules could be defined that form the basis for the development of algorithms in a computer system. The system enables an optimization of online advertising campaigns (and the related costs) with the consideration of the basic business targets regarding the sales of services. The solution presented here was tested in real world processes and it proved to be significantly useful.