

---

## Profiling and Segmenting Clients with the Use of Machine Learning Algorithms

---

Submitted 10/03/21, 1st revision 13/04/21, 2nd revision 10/05/21, accepted 30/06/21

Paweł Rymarczyk<sup>1</sup>, Piotr Gołabek<sup>2</sup>, Sylwia Skrzypek - Ahmed<sup>3</sup>,  
Magdalena Rzemieniak<sup>4</sup>

### **Abstract:**

**Purpose:** The aim of the article is to develop a solution for customer profiling and segmentation using modern machine learning methods.

**Design/Methodology/Approach:** Models were developed to improve the analysis of data, human behavior, data mining business processes, and as a result, the creation and provision of new improved solutions using machine learning algorithms. The GRU method was used, which is a simplified but also a more streamlined version of the LSTM cell offering similar performance with a much lower computation time.

**Findings:** The main purpose of the developed solution is to enable and improve the analysis of profiling and segmentation of customers for forecasting sales, due to the possibility of detecting or determining additional seasonal effects.

**Practical Implications:** Effective tools have been developed to enable customer segmentation. A more complex model was used, taking into account the sale, especially in the sense of the time series in which the sale took place. In its form, the model consists of a trend function modeling non-periodic changes in the value of time series periodic changes.

**Originality/Value:** A novelty is the use of the GRU network, which is an improved version of the standard recursive neural network and a simplified version of the standard LSTM network. Similarly to LSTM networks, it aims to solve the problem of a vanishing gradient, i.e., its disappearance or explosion. In the presented solution, a more complex model was used, consisting of several components and taking into account sales, especially in the sense of the time series in which the sale took place.

**Keywords:** Machine learning, forecasting, data mining, LSTM.

**JEL codes:** C45, C53, M30.

**Paper type:** Research article.

---

<sup>1</sup>Corresponding Author: Netrix Group sp. z o.o. , Lublin, Poland;  
e-mail: [pawel.rymarczyk@netrix.com.pl](mailto:pawel.rymarczyk@netrix.com.pl)

<sup>2</sup>University of Economics and Innovation in Lublin, Lublin, Poland;  
e-mail: [piotr.golabek@wsei.lublin.pl](mailto:piotr.golabek@wsei.lublin.pl)

<sup>3</sup>Same as in 2, e-mail: [a.Skrzypek-Ahmed@wsei.lublin.pl](mailto:a.Skrzypek-Ahmed@wsei.lublin.pl)

<sup>4</sup>Faculty of Management, Lublin University of Technology, Lublin, Poland,  
e-mail: [m.rzemieniak@pollub.pl](mailto:m.rzemieniak@pollub.pl)

## 1. Introduction

The customer profile on the market is constantly changing. This contributes to changing the business context and the conditions in which enterprises compete. Buyers are getting weary of the abundance of offers and less and less loyal to the brand. They perfectly know the opinions of other people about products and services, and they constantly increase their requirements (Bachnik, 2010; Bojanowska and Kulisz, 2020). Many authors indicate that the study of consumer behavior is important because it shapes the attitudes and preferences of the customer, allows to determine its value for the company and to forecast and understand the demand for products (Antonides and van Raaij, 2003). The future behavior of the buyer is significantly influenced by his satisfaction as a consumer (Anderson and Sullivan, 1993), i.e., *"emotional reaction to comparative processes carried out by the client, consisting in comparing one's experiences and sensations after consuming a product or service with expectations"* (Tarczydło, 2011). Measurement of satisfaction, loyalty, satisfaction and customer value generate indicators critical for business results (such as profit or market share) (Skowron *et al.*, 2020). Therefore, the condition for a successful prospering of the company is meeting customer expectations (Hallencrutz and Parmler, 2019), i.e., adjusting products to their expectations, caring for relationships with them both before and after the transaction (Drucker, 1954), as well as making every effort to predict its needs earlier than the competition (Bachnik, 2010).

Customer segmentation is the process of dividing heterogeneous customers into homogeneous groups based on common attributes and is essential to more effectively serve different customers with rich sets of diverse customer preferences. Customer segmentation studies used at least one of the following data types (Cui *et al.*, 2006):

- a) Demographic data, e.g., gender, age, marital status, household size etc.
- b) Geographic data, e.g., place of residence or work, etc.
- c) Psychographic data, e.g., social class, lifestyle, personality traits etc.
- d) Attitude data, e.g., perceived data gathered from surveys that capture information about what people say they are doing to understand and interpret shopper behavior (Konucs, Verhoef, and Neslin, 2008).
- e) Sales data showing purchasing behavior, e.g., sales volume, number of visits, frequency of visits, cash amount, days between journeys, time of purchase, last visit, etc., and customer needs and preferences according to the combination of buyer categories.
- f) Behavioral data, i.e., data indicating non-shopping behavior, e.g., data from RFID-enabled carts that show what buyers are putting in their basket, data from an RFID-enabled fitting room, which clothes people are trying on, navigation data from a website or storing data from Bluetooth Low Energy (BLE) technology (Skogster and Uotila, 2008).

By combining data from sales systems and behavioral data related to the movement of customers in the area of the sales space, it is possible to build systems that allow you to optimize orders, how to arrange goods and other customer behavior patterns.

The main purpose of the developed solution is to enable and improve the analysis of multi-source data, human behavior, data mining business processes, and as a result, the creation and provision of new improved procedures and solutions using artificial intelligence algorithms. In today's complex business world, companies need to pioneer ways to differentiate themselves from other market participants by becoming more cooperative, effective, precise, and flexible. They must be able to react quickly to market needs and changes. Depending on the company's competitive advantage, which may be novelty, price, great website content or social media presence, specific online strategies must be applied to reach the desired market. It's worth noting that the data we store and somehow use it can build a market advantage. Data and information are becoming essential resources for many organizations. The advent of new technologies and data collection tools is now opening up areas of research into understanding customer-business relationships and interactions. Conventional user interest modeling methods are not developed for integration and interest research from multiple sources, therefore they are not very effective in obtaining a relatively complete description of user interests in a distributed setting.

The issues considered during the analyzes are, identifying different groups of customers visiting stationary and online stores, understanding the specific needs and preferences of each segment, offering appropriate services to meet customer needs (Griva *et al.*, 2018). The researchers' response to the needs of the retail industry was to develop effective tools for customer segmentation. There have been many studies using different types of data.

## 2. Forecasting with Machine Learning

Before machine learning can be used, the problems with predicting time series must be framed in supervised machine learning. When  $H$  historical data record is available, the problem of one-step forecasting can be solved as a supervised machine learning problem. Supervised learning models, from a finite set of  $X$  observations, the relationship between the set of input fields and one or more output fields that are believed to be dependent on the input data. Once a mapping model is available, it can be used for one-step forecasting. In one-step forecasting,  $n$  previous series values are available, and the forecasting problem can be presented as a general regression problem (Bontempi, Ben Taieb, and Le Borgne, 2012).

In the prediction setting, the training set is derived from the historical series  $H$  by creating the input matrix  $X = [(N - n - 1) \times n]$  and the output vector  $Y = [(N - n - 1) \times 1]$ .

$$X = \begin{bmatrix} y_{N-1} & y_{N-2} & \dots & y_{N-n-1} \\ y_{N-2} & y_{N-3} & \dots & y_{N-n-2} \\ \vdots & \vdots & \vdots & \vdots \\ y_n & y_{n-1} & \dots & y_1 \end{bmatrix} \tag{1}$$

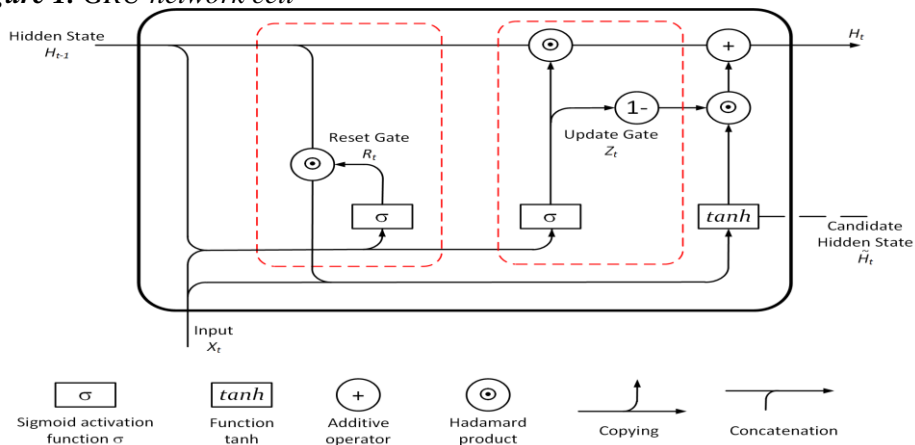
GRU (Gated Recurrent Unit) networks are somewhat of an improved version of the standard recursive neural network and a simplified version of the standard LSTM network. Similarly to LSTM networks, they are designed to solve the problem of a vanishing gradient, i.e. its disappearance or explosion. Therefore, it is worth considering the consequences of ignoring such a problem:

- It is very common to encounter a situation where one of the initial observations is very important for the task of predicting all subsequent observations. For such cases, it is important to have mechanisms in place to store such information in a memory cell. Without such a mechanism, a very large gradient should be assigned to the exemplary observation, due to the fact that it has a large impact on all subsequent observations.
- Another common situation is that some tokens do not give us any relevant information. In this case, you would need a mechanism to omit such tokens in the hidden state representation.
- Another situation is when there is some logical gap between successive parts of the sequence, so it would be important to be able to reset the internal state representation at the appropriate point in time.

### 2.1 GRU Method

The GRU method is a simplified but also a more streamlined version of the LSTM cell, which offers similar performance with much less computation time. The GRU network model illustrating the repeating network unit is shown in Figure 1.

Figure 1. GRU network cell



Source: Own creation.

To solve the vanishing gradient problem found in standard RNNs, the GRU uses the so-called update gate and reset gate as shown in the figure above. Basically, these are the two vectors that decide what information should be passed to the output. Their peculiarity is that they can be trained to preserve historical information without having to transform it over time or delete information that is irrelevant to the predictions. The core and most important part of the GRU model are reset and update gates. Their construction is based on the principle that they are vectors with values in the interval (0,1) so that it is possible to perform convex operations. After such adaptation of the network, the reset gate would allow, for example, to control what part of the previous state we want to still remember and keep. In a similar way, the updating gate allows us to control what part of the new state is a copy of the previous state.

The figure below shows the input stage for both the reset and updating gate, taking into account the input for the current time step and the hidden state from the previous time step. The outputs of both gates are given by two combined layers using a sigmoidal activation function that transforms the input data contained in  $\mathbb{R}$  to the interval (0,1) (Chung *et al.*, 2014):

$$\text{sigmoid}(x) = \frac{1}{1 + \exp(-x)}. \quad (2)$$

In mathematical terms, let's assume (as shown in the figure) that for a certain time step  $t$ , the input vector is marked with  $X_t \in \mathbb{R}^{n \times d}$ , where  $n$  - number of observations,  $d$  - number of inputs, and the hidden layer from the previous time step is marked with is given by  $H_{t-1} \in \mathbb{R}^{n \times h}$ , where  $h$  denotes the number of hidden units. Then the reset gate  $R_t \in \mathbb{R}^{n \times h}$  and the updating gate  $Z_t \in \mathbb{R}^{n \times h}$  are determined as follows:

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r), \quad (3)$$

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z), \quad (4)$$

where  $W_{xr}, W_{xz} \in \mathbb{R}^{d \times h}$  and  $W_{hr}, W_{hz} \in \mathbb{R}^{h \times h}$  represent the weights, and  $b_r, b_z \in \mathbb{R}^{1 \times h}$  the deviations. To transform the values and bring them to the interval (0,1) we use the sigmoid function mentioned earlier.

In the next step, you should integrate the character obtained by the reset gate with the general form of the hidden state update for RNN, i.e., the form:

$$H_t = \phi(X_t W_{xh} + H_{t-1} W_{hh} + b_h) \quad (5)$$

with the previously adopted assumptions and for the appropriate function  $\Phi$ . Such a procedure leads to the determination of the candidate for the hidden layer  $\tilde{H}_t \in \mathbb{R}^{n \times h}$  in the time step  $t$  as:

$$\tilde{H}_t = \tanh(X_t W_{xh} + (R_t \odot H_{t-1}) W_{hh} + b_h), \quad (6)$$

where  $W_{xh} \in \mathbb{R}^{d \times h}$ ,  $W_{hh} \in \mathbb{R}^{h \times h}$ ,  $b_h \in \mathbb{R}^{1 \times h}$  and the symbol  $\odot$  denotes the Hadamard product. The introduced tanh function also ensures that the values in this hidden state will be in the range (-1,1). The described state is commonly referred to as "candidate" due to the fact that the update function should be considered for it next.

To obtain the form presented in Fig. 1, it was only necessary to take into account the operation of the updating gate  $Z_t$ . On its basis, it is possible to determine the degree to which the new hidden state  $H_t \in \mathbb{R}^{n \times h}$  consists of the previous hidden state  $H_{t-1}$  and the degree to which the newly created candidate state  $\tilde{H}_t$  is used. For such a need, the updating gate  $Z_t$  can be used, assuming only convex combinations between the states  $H_{t-1}$  and  $\tilde{H}_t$ . Therefore, the final updating equation for the GRU model is as follows:

$$H_t = Z_t \odot H_{t-1} + (1 - Z_t) \odot \tilde{H}_t. \quad (7)$$

Thus, each time the value for the updating gate comes close to 1, the previous state is kept. Also note that information derived from  $X_t$  is generally ignored, effectively skipping step  $t$  in the dependency chain. In the opposite case, ie when  $Z_t$  is close to 0, the new hidden state  $H_t$  converges to the candidate  $\tilde{H}_t$ . Such a model can effectively deal with the decaying gradient problem in the RNN and better emphasize the relationships for the sequences in longer time steps, as it does not clean up new input data each time, but retains the relevant information and passes it on to the next time steps of the network. Such a carefully trained model can cope very well even in very complex situations.

## 2.2 Construction of an Artificial Neural Network

The tool used to build a network using GRU cells is the Keras environment based on TensorFlow. The model has a sequential structure with two layers with GRU cells and one dense layer as the output layer. The selection of optimal hyperparameter values was carried out using the Talos environment (Autonomio, 2019). The dictionary of parameters and their values that were selected are presented in Table 1.

The Talos library combines various hyperparameter optimization strategies with an emphasis on maximizing flexibility, efficiency and the results of the adopted random strategy. It works with the Keras library used in this analysis and provides a fully automated Prepare  $\rightarrow$  Optimize  $\rightarrow$  Deploy flow that handles forecasting issues.

**Table 1.** A set of hyperparameters tested in the learning process

'units_1'	5, 20
'units_2'	5, 10
'batch_size'	72, 36
'epochs'	100
'dropout'	0.05, 0.1
'lr'	0.1, 0.01, 0.001, 0.0001
'optimizer'	Adam, Nadam, RMSprop, Adamax
'losses'	'mean_absolute_error', 'mean_squared_error'
'last_activation'	'sigmoid'

*Source: Own creation.*

### 2.3 Learning Process

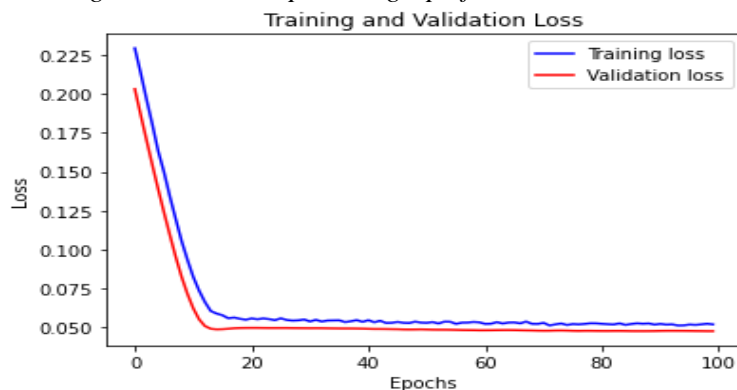
During the learning process, along with searching the hyperparameter matrix, two models were selected, the settings and learning charts of which are presented below.

**Table 2.** The set of hyperparameters defined for Model 1

units_1	20
units_2	10
batch_size	72
epochs	100
dropout	0.05
optimizer	Adamax
learning_rate	0.1
last_activation	sigmoid
loss measure	mean_absolute_error

*Source: Own creation.*

A graphical representation of the learning and verification process of model 1 is shown in Figure 2.

**Figure 2.** Learning and validation process graph for Model 1

*Source: Own creation.*

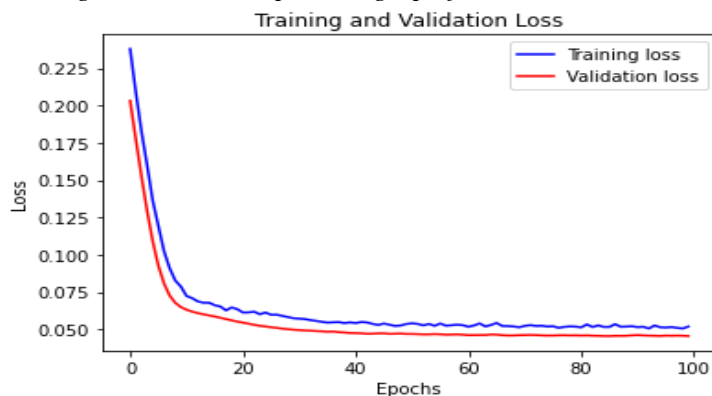
The experiments related to training the model were carried out in the form of several hundred attempts to search the hyperparameter matrix with the Talos library. Based on the conducted experiments, another model was selected 2.

**Table 3.** The set of hyperparameters defined for Model 2

units_1	20
units_2	5
batch_size	36
epochs	100
dropout	0.05
optimizer	Adam
learning_rate	0.1
last_activation	sigmoid
loss measure	mean_absolute_error

*Source:* Own creation.

**Figure 3.** Learning and validation process graph for Model 2



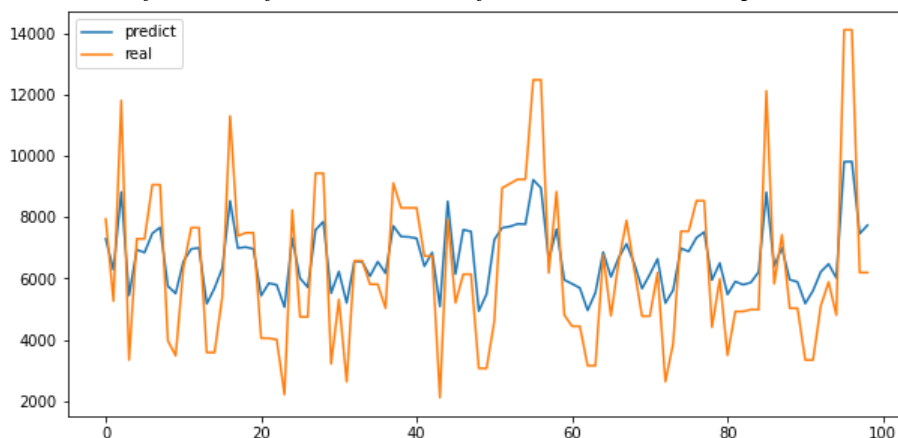
*Source:* Own creation.

If the model is well trained, its performance should be similar in both the training and validation set. In both figures above, the learning and validation errors stabilize after 20 epochs. The plotted losses are normalized values. The plots shown (Figure 2, Figure 3) show that the networks have not been overtrained and provide a good fit to the data.

### 2.4 Sales Forecasts - Model Verification

Then, on the basis of the model, a prediction was carried out on the basis of previously determined test data, the results of which are presented in Figure 4. Thus, you can see a clear good prediction of the model in terms of the trend of jumps or falls in sales. In order to return to the initial values, the values had to be scaled with the inverse function to the previously described MinMax function. The RMSE for the obtained prediction was 1362.67.



**Figure 4.** Plot of real and forecasted values for the Sales\_In\_Groups variable

*Source:* Own creation.

### 3. Conclusions

Sales forecasting allows companies to act well in advance, among other things, it supports decisions in the field of planning marketing activities and determines the time of introducing and withdrawing individual products from the market. The result of forecasting may be the sum of sales with a breakdown into product groups or distribution channels. If there is a relationship between sales and the number of products or points of sale, we will use the linear regression method to calculate the expected sales volume, which will allow us to forecast in the simplest way. On the other hand, when the determination of the sales volume depends on a greater number of variables, we will use the multiple regression method for analysis.

In the presented solution, a more complex model was used, consisting of several components and taking into account sales, especially in the meaning of the time series in which the sale took place. In its form, the model consists of a trend function modeling non-periodic changes in the value of the time series, periodic changes (such as weekly or annual seasonality), holiday effects - usually occurring in the same time intervals, e.g. annually, and of course the error about which it is assumed that it has a normal distribution. Sales forecasting gives very accurate results due to the possibility of detecting or determining additional seasonal effects.

Forecasting affects the activities of all areas of the company, supports decision-making in determining the amount of purchases, production planning, logistics, sales implementation, as well as marketing campaigns and after-sales services. If the forecasts significantly exceed the actual performance, then there may be an unnecessary freezing of cash, an increase in storage costs, a reduction in prices and business profitability. Underestimating forecasts, on the other hand, affects the shortage of goods in warehouses, delays in logistics, untimely service, which in turn causes customer dissatisfaction.

**References:**

- Anderson, E.W., Sullivan, M.W. 1993. The Antecedents and Consequences of Customer Satisfaction for Firms. *Marketing Science*, 12(2), 125-143. <https://doi.org/10.1287/mksc.12.2.125>.
- Antonides, G., van Raaij, W.F. 2003. *Zachowanie konsumenta. Podrecznik akademicki*. Warszawa: Wydawnictwo Naukowe PWN.
- Autonomio, 2019. Talos. <http://github.com/autonomio/talo>.
- Bachnik, K. 2010. Wybrane trendy w zarządzaniu w dobie kryzysu. In *Studia i Prace Kolegium Zarządzania i Finansów*, 15-21. Warszawa: Szkoła Główna Handlowa.
- Bojanowska, A., Kulisz, M. 2020. Polish Consumers' Response to Social Media EcoMarketing Techniques, *Sustainability*, 12(21), 8925. <https://doi.org/10.3390/su12218925>.
- Bontempi, G., Ben Taieb, S., Le Borgne, Y.A. 2013. Machine Learning Strategies for Time Series Forecasting. In: Afaure, M.A. and Zimányi, E. (Eds.) *Business Intelligence. eBISS 2012. Lecture Notes in Business Information Processing*, vol 138, 62-77. Berlin, Heidelberg: Springer. [https://doi.org/10.1007/978-3-642-36318-4\\_3](https://doi.org/10.1007/978-3-642-36318-4_3).
- Chung, J., Gulcehre, C., Cho, K., Bengio, Y. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv Prepr. arXiv1412.3555*.
- Cui, G., Wong, M.L., Lui, H.K. 2006. Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science*, 52(4), 597-612. <https://doi.org/10.1287/mnsc.1060.0514>.
- Drucker, P.F. 1954. *Management by Objectives and Self-Control*. The Oxford Handbook of Management.
- Griva, A., Bardaki, C., Pramatari, K., Papakiriakopoulos, D. 2018. Retail business analytics: Customer visit segmentation using market basket data. *Expert Systems with Applications*, 100, 1-169. <https://doi.org/10.1016/j.eswa.2018.01.029>.
- Hallencreutz, J., Parmler, J. 2019. Important drivers for customer satisfaction--from product focus to image and service quality. *Total Quality Management & Business Excellence*, 32(5-6), 501-510. <https://doi.org/10.1080/14783363.2019.1594756>.
- Konucs, U., Verhoef, P.C., Neslin, S.A. 2008. Multichannel shopper segments and their covariates. *Journal of Retailing*, 84(4), 398-413.
- Skogster, P., Uotila, V. 2008. Collecting consumer behavior data with WLAN. *International Journal of Information Systems and Supply Chain Management (IJSSCM)*, 1(2), 57-75. <https://doi.org/10.4018/jisscm.2008040104>.
- Skowron, Ł., Gąsior, M., Sak-Skowron, M. 2020. The Impact of a Time Gap on the Process of Building a Sustainable Relationship between Employee and Customer Satisfaction. *Sustainability*, 12(18), 7446. <https://doi.org/10.3390/su12187446>.
- Tarczydło, B. 2011. Badanie satysfakcji klientów-studium przypadku. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Ekonomiczne Problemy Usług*, 2011, 117-126.