

Churn Detection and Prediction in Automotive Supply Industry

Hasan Can Karapinar
Istanbul Technical University,
Faculty of Management, 34357,
Macka, Istanbul, Turkey
Email:
hasancankarapinar@gmail.com

Ayca Altay
Istanbul Technical University,
Faculty of Management, A304,
34357, Macka, Istanbul, Turkey
Email: altaya@itu.edu.tr

Gulgun Kayakutlu
Istanbul Technical University,
Faculty of Management, B309,
34357, Macka, Istanbul, Turkey
Email: kayakutlu@itu.edu.tr

□ *Abstract*—Companies have both large certified enterprises and small unauthorized service providers as their competitors in the automotive supply industry. As technology related industries undergo more intensive competition, churn detection and prediction become essential to be precautious about leaving customers. The literature for churn detection offers numerous statistical and intelligent methods. In this study, Artificial Neural Networks and Decision Trees are applied to detect the churn in and analyze the validity of these methods for the automotive supply industry. The problem involves both categorical and continuous numerical decision inputs which cannot simultaneously fed into Decision Trees. In this case, continuous inputs should be divided into binary categorical ones by splitting into various intervals which are called buckets. Particle Swarm Optimization algorithm is implemented for finding optimal buckets for the churn problem data. Results indicate that while both algorithms are promising, the bucket tuning for Decision Trees complicate the churn detection process.

I. INTRODUCTION

Retaining existing customers has become the main objective of Customer Relationship Management (CRM) departments in markets with fierce competition [1]. Automobile service industry stands for a real example when fierce competition is mentioned in Turkey. Maintenance and hard repairs are the most profitable cases in automobile service market, and every one of the authorized and non-authorized services are trying to take a big pie of this market. Developed subsidiary industry of automobile spare parts and low labor costs of non-authorized services are challenging automobile companies to retain their customers, making churn prediction an important part of this challenge. Retaining a customer is more profitable than acquiring a new one so that companies are willing to know which customer will churn like in other markets. Preventive actions, new marketing strategies and long term service contracts can be made in case of foreknow the churners. Churn prediction must be as accurate as possible to ease decision making. It is known that the cost of earning new customers is a lot more than sustaining existing ones. Hence, it becomes essential to

select the best churn detection methods given company data [2].

In literature, there are various methods for data mining in churn prediction; yet, the common point of these methods are that they yield a classification and/or clustering decision where the customers are grouped as "churned" and "not churned" [3]. The methods applied for grouping customers for churn can mainly be subsumed under statistical, intelligent and hybrid methods. Statistical methods involve approaches that intend to minimize incorrect classification using statistical data processing tools such as Decision Trees or Regression [4]. Learning algorithms and intelligent methods make up the recent trend with the contribution of computer technologies in churn mining where methods such as Neural Networks and metaheuristics are implemented [5]. Lastly, hybrid methods that combine statistical and learning methods are stated to provide more accurate results [5]. These methods will be elaborated in the next section.

The literature of churn prediction is intensified in the service industry; especially in telecommunication and banking [3, 6-8]. Emerging from these industries, churn prediction is now known to contribute to companies in other industries, as well. In this study, we have applied two different data mining techniques on the problem of churning customers in the automotive supply industry. Artificial Neural Network and a hybrid method (Particle Swarm Optimization (PSO) based ID3 Decision Trees) are used for determining the churn and their accuracies were compared in order to provide a better detection performance.

The structure of this paper is as follows: The next section provides a literature review on churn determination and prediction together with the motivation of this study. The third section introduces the methodology and an application is provided in the following section. Finally, the conclusions are presented in the last section.

II. LITERATURE REVIEW

Churn analysis and prediction aims to detect churn customers and/or predicting how likely the customers are to be churned [3]. The literature of churn analysis is intensified in industries where high rate of competition exists; such as

□ This work was not supported by any organization

the telecommunication industry where the churn rate reaches its peak with an annual average of 27% [9].

The implementation of data mining techniques in churn analysis is inevitable, since the extraction of patterns in churned customers is essential for determining and predicting churn [10]. An analytic view of the methods in churn analysis can be analyzed three-fold: statistical, intelligent and hybrid methods. The most exploited method in statistical churn analysis is Decision Trees and Random Forests [4]. Ahn et al. [11] use Factor Analysis in telecommunication industry and state that customer status is the most important criteria. Bin et al. [12] implement Decision Trees in phone sales. Wei and Chiu [3] implement decision trees that extract rules with C4.5 algorithm for churn prediction. Larivière and Van del Peol [13] implement logistic and linear regression models and extend the Decision Tree approach and execute random forests in telecommunication industry. The implementation of decision trees is also diversified. For instance, Lemmens and Croux [14] use Bagged & Boosted Decision Trees in Telecommunications Industry.

The execution of intelligent and learning methods have been the focus of literature of the past two decades. In particular, Artificial Neural Networks (ANNs) methods, both supervised and unsupervised, represent the majority of studies in this branch. Hadden et al. [5] implement basic Back Propagation algorithm on Feed-Forward Neural Networks. Buckinx and Van den Poel [15] compare the performance of Random Forests and ANNs and conclude that the prediction by ANNs is more robust in case of FMCG industry example. Other variations of ANNs involve Support Vector Machines (SVMs) and Self Organizing Maps (SOMs). Coussemont and Van den Poel [16] construct and SVM model to predict churn customers in telecommunications industry, whereas Tsai and Lu conduct a similar study where they implement an unsupervised approach by SOMs [17]. Even though, ANNs are known to be stronger churn predictors than Decision Trees, they are known to have disadvantages such as early convergence or being stuck at local optima [5]. Hence, these models are enhanced with other metaheuristics or Artificial Intelligence methods. For example, Karahoca and Karahoca [18] implement an Adaptive Neuro Fuzzy Inference System model integrated with a fuzzy clustering method which highly outperforms ANNs.

Hybrid methods are known to provide better comprehension of the churn structure. Xia and Jin [19] first find out the criteria that affect churn using Factor Analysis and using these criteria, they build an SVM model in order to predict churn. Idris et al. [9] find the optimum decision tree using Particle Swarm Optimization and conclude that metaheuristic-based decision trees outperform in terms of classification accuracy.

In this study, one main statistical method - Decision Trees - and one main intelligent method - ANNs - are tested on automotive supply chain determination. The very basic ID3

algorithm is used for Decision Trees. However, the ID3 (Interactive Dichotomizer 3 (ID3) algorithm for Decision Trees only operate when all variables are categorical and the data on hand involves continuous numerical variables. Hence a separation scheme is mandatory. The optimum separation points are found by Particle Swarm Optimization algorithm, yielding a hybrid method. A comparison of a hybrid and an intelligent method is constructed in terms of automotive supply industry perspective.

The factors used for classification in literature are observed to depend industry-wise. Telecommunication industry related studies - GSM churn, to be concise - are mainly focused on minutes of usage, monthly number of calls or monthly bills, which are specific to this industry [3]. Hence, it is essential to find out the criteria that are specific to the industry. However, the literature on automotive supply industry does not offer many criteria; which is why, mainly interviews with industry experts have determined the criteria for this study.

III. METHODOLOGY

Decision Trees and ANNs will be compared for this study, since they are the basic decision making.

A. Decision Trees

The basic ID3 Algorithm is used for Decision Trees. Proposed by Quinlan [20], the algorithm uses information gain in order to decide the splits or branches of the tree, which is calculated as given below:

$$IG(S) = \sum_{k=1}^N -p(k) \log_2 p(k) \quad (1)$$

where S is an element of the input attribute set I , N is the number of unique attribute values, $p(k)$ is the probability of this attribute value k [20-21].

For each node, the attribute that provides the most information gain is split until final branches (called leaves) are reached. The algorithm builds a decision tree from the data which are discrete in nature. The data in the case of automotive supply industry churn involves continuous variables such as mileage. In this case, the data of continuous attributes should be discretized. One way to handle such continuity problems is to assign separation or split points for the data and convert them into binary variables. This process is called "splitting into buckets" where each bucket is a class that specifies a range of values and a data point is subsumed under one of these ranges or buckets [22]. For example, if the range of a continuous variable is $[0,10]$ and the separation point is 5, then, the buckets are $[0,5]$ and $[5,10]$. A value, say 3, is classified in a binary fashion as 1 and 0, for two buckets, respectively. Moreover, the selection of separation points affects the structure of the input data, and hence, largely impacts the accuracy of the decision tree. Therefore, these separation points should be tuned. This tuning is achieved by the PSO algorithm.

B. PSO Algorithm

The PSO algorithm offered by Kennedy and Eberhart [23]. The steps of the conventional PSO algorithm are given below:

Let p_i be the i^{th} particle in the swarm which consists of N particles and let each particle have n variables.

Step 1. The particle velocities and positions are initiated as priorities

$$x_{i,j} = x_{min} + r(x_{max} - x_{min}), \quad i = 1, \dots, N, j = 1, \dots, n \quad (2)$$

$$v_{i,j} = \alpha \frac{x_{min} + r(x_{max} - x_{min})}{\Delta t} \quad i = 1, \dots, N, j = 1, \dots, n \quad (3)$$

where x denotes the position, v denotes the velocity and is constant in the range $[0,1]$.

Step 2. The objective function values of particles are calculated as $f(x_i)$.

Step 3. The best position for each particle and the global best position for the swarm are updated.

$$\text{If } f(x_i) < f(x_i^{pb}) \text{ then } x_i^{pb} \leftarrow x_i \quad (4)$$

$$\text{If } f(x_i) < f(x_i^{sb}) \text{ then } x_i^{sb} \leftarrow x_i \quad (5)$$

where pb denotes the particle best and sb denotes the swarm best.

Step 4. Particle velocity and particle position are updated, that is, the new velocities and positions are calculated for each particle.

$$v_{i,j} \leftarrow wv_{i,j} + c_1r_1 \left(\frac{x_i^{pb} - x_{i,j}}{\Delta t} \right) + c_2r_2 \left(\frac{x_i^{sb} - x_{i,j}}{\Delta t} \right), \quad i = 1, \dots, n \quad (6)$$

where $wv_{i,j}$ is the separation term, $c_1r_1 \left(\frac{x_i^{pb} - x_{i,j}}{\Delta t} \right)$ is the alignment term and $c_2r_2 \left(\frac{x_i^{sb} - x_{i,j}}{\Delta t} \right)$ is the cohesion term.

Step 5. Step 2 is returned until a termination criterion is satisfied.

C. ANNs

ANN models are basically derived from nervous systems which are able to perform functional input-output mapping such as machine learning and pattern recognition [24-25]. In this study, feed forward neural networks with back-propagation learning algorithm are utilized.

ANNs are constructed by layers of cells which are called neurons as can be seen in Figure 2. Neurons connect to other neurons in consecutive layers by a certain weight value. Inputs of each neuron are the weighted sum of the weights of incoming neural connections and a bias. This sum is subjected to a transfer or an activation function within the neuron [23]. The input of a neuron is calculated as

$$n = \sum_i w_i \cdot x_i + bias \quad (7)$$

where n is the input of a neuron, w_i is the weight of the i^{th} neuron in the previous layer and x_i is i^{th} input value.

The Back-propagation algorithm is a learning structure that minimizes the forecasted output of the network and the actual output by adjusting weights of the network. In order to achieve this error minimization, the data are first fed into the network and the error term obtained is propagated back into the network to adjust weights with formula given below

$$\Delta w_{ji}^k(n+1) = -\eta \frac{\partial E}{\partial w_{ji}^k} + \Delta w_{ji}^k(n) \quad (8)$$

where n is the index for iterations, η is the learning rate and E is the error term. Note that the derivative of the error term with respect to the weights yields the weight adjustment.

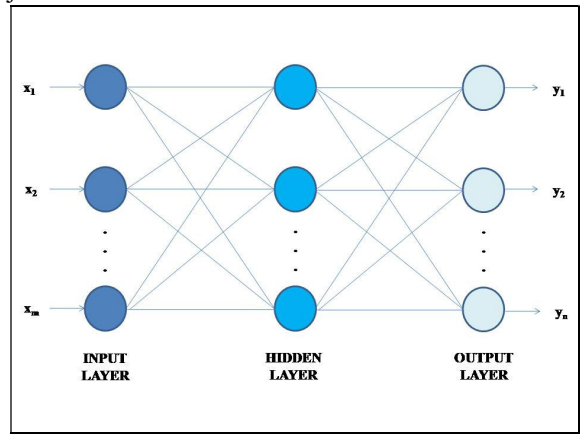


Fig. 2 The structure of an ANN

IV. APPLICATION AND RESULTS

In the last decade, there has been an increase in fleet customers in automotive market. Instead of buying a new car, companies started to rent fleet of cars. So that, detecting a fleet car has become critical because detecting one of the fleet car's churning means losing all of the cars and profit from that fleet.

In this study, the churn determination of customers of a leading car seller of Turkey is achieved using a PSO-based ID3 Decision Tree and ANNs. The criteria for churn evaluation are extracted from the interviews with industry experts and company's academic consultants.

The criteria for churn evaluation are listed below:

1. Annual penetration of 2013: It represents the total amount of cash inflow that customer did throughout the year 2013. It is a continuous attribute; hence, it should be bucketed.
2. Annual penetration of 2014: It represents the total amount of cash inflow that customer did throughout the year 2014. It is a continuous attribute and it should be bucketed as in the previous year's case.

3. Annual penetration of 2015: It represents the total amount of cash inflow that customer did throughout the year 2015. Just like the previous years, it is bucketed.
4. Annual frequency of maintenance of 2013: It represents the number of times that the car was taken to maintenance services throughout the year 2013. It is a discrete variable.
5. Annual frequency of maintenance of 2014: It represents the number of times that the car was taken to maintenance services throughout the year 2014. It is a discrete variable.
6. Annual frequency of maintenance of 2015: It represents the number of times that the car was taken to maintenance services throughout the year 2015. It is a discrete variable.
7. Year of purchase: It represents the year that the car was purchased by its owner. It is a discrete variable.
8. Insurance: It represents if the car is insured. It is a binary discrete variable.

Discretization of the continuous variables is achieved by splitting into buckets for the first three criteria; hence, PSO algorithm should be applied in order to find the optimum separation points. However, the optimum number of buckets is not known a priori; and thus, number of buckets are found by trial and error. The splits for bucket construction of 2, 3, 4 and 5 buckets are tuned by the PSO algorithm.

The performance measure (objective function) of the algorithm is taken as the rate of misclassifications. There are a total of 500 data points and their churn conditions are known. Two sample data points are shown in Fig. 3. The first customer is a churn and the second customer has not

TABLE I.
ALGORITHM RESULTS FOR TRAINING DATA

Method	Average misclassification Percentage (μ)	Standard deviation of the misclassification percentage (δ)
PSO – based Decision Tree with 2 buckets (PSODT2)	0.086	0.12
PSO – based Decision Tree with 3 buckets (PSODT3)	0.118	0.05

PSO – based Decision Tree with 4 buckets (PSODT4)	0.150	0.07
PSO – based Decision Tree with 5 buckets (PSODT5)	0.114	0.04
ANNs	0.109	0.02

TABLE III.
ALGORITHM RESULTS FOR TESTING DATA

Method	Average misclassification Percentage (μ)	Standard deviation of the misclassification percentage (δ)
PSODT2	0.154	0.05
PSODT3	0.170	0.03
PSODT4	0.167	0.05
PSODT5	0.198	0.04
ANNs	0.160	0.03

churned. 70% of the data are used for training and the rest is used for testing. PSO parameters are adjusted as 20 for the swarm size, 1 for the inertia coefficient (w), 2 for the cognitive coefficient (c_1) and 2 for the social coefficient (c_2). The learning rate (η) of the ANN is 0.7. Table 1 shows the average and the standard deviation of 30 runs of training of the PSO-hybrid and ANN results, respectively. Table 2 shows the same results for the testing data.

PSO-based Decision Trees yield the best result when the number of buckets is 2 for the training data and the testing data. Additionally, the PSO-hybrid with 2 buckets has one run without any error for the training data; whereas the ANN has one run without any error for testing data. In order to analyze if the results of these 30 runs are statistically significant, t tests among the algorithm results is applied. Tables 3 and 4 display the p values of the t test results of methods in a pairwise comparison fashion. The lower parts of the tables are not filled because the tables are symmetrical. For example, the value that is in the intersection cell of PSODT2 and PSODT3 also applies to the value of the intersection cell of PSODT3 and PSODT2.

TABLE II.
SAMPLE DATA POINTS

Rack Number	Annual Penetration			Annual Maintenance Frequency			Year of purchase	Insurance	Churn
	2013	2014	2015	2013	2014	2015			
XXXXXXXXXX	0	478,546	647,431	0	1	1	2013	1	1
YYYYYYYYYY	0	12,37	0	0	1	0	2012	0	0

TABLE IV.
P VALUES OF T TESTS FOR TRAINING DATA

	PSODT2	PSODT3	PSODT4	PSODT5	ANNS
PSODT2	-	0.2113	0.0640	0.7335	0.3047
PSODT3		-	0.0340	0.8648	0.4793
PSODT4			-	0.0175	0.0031
PSODT5				-	0.5427
ANNS					-

TABLE III.
P VALUES OF T TESTS FOR TESTING DATA

	PSODT2	PSODT3	PSODT4	PSODT5	ANNS
PSODT2	-	0.1383	0.3181	0.0004	0.5752
PSODT3		-	0.7791	0.0330	0.2018
PSODT4			-	0.0103	0.5134
PSODT5				-	0.0001
ANNS					-

The less the p value is, the more statistically different the results of the two methods compared. For example, for the training data, the p value between PSODT2 and PSODT3 is 0.0640. If a significance level of 10% is assigned, the given p value is less than 0.1, that is, the results of PSODT2 and PSODT3 are statistically significantly different. Moreover, from Table 1, it can be seen that the misclassification error rate of PSODT3 is larger, which means that PSODT2 provides better results than PSODT3. According to Table 3, for the training data, apart from PSODT4, other algorithm results are not statistically significantly different from each other. In terms of the testing data (see Table 4), PSODT5 is the worst result providing algorithms and other algorithm results are similar. Using these results, the non-dominated algorithms are PSODT2, PSODT3 and ANNs. ANNs provide better results for both training and testing data. Likely, the results of PSODT2 and PSODT3 are also as good; yet, in terms of Decision Trees, the bucket numbers have to be optimized. Moreover, considering the accuracy of the models, it is concluded that the criteria determined by industry experts are valid for churn detection.

V. CONCLUSION

Churn detection is difficult in industries where the competition is fierce. The churn is intensive in especially technology related industries. Even though not being as technologically intensive as online markets or telecommunication industries, automotive supply industry is also one of the industries that suffers from frequent churn. Therefore, churn detection and prediction becomes essential.

In this study, ANNs are constructed for churn prediction for automotive supply industry. Traditional churn approaches also offer the implementation of Decision Trees to be a powerful means of churn analysis. However, in the

presence of non-categorical (numeric – continuous) data, the implementation of Decision Trees with raw data may be misleading. For this reason, it is obligatory that the continuous data are converted to binary and categorical data. This process is achieved by using ‘buckets’. PSO algorithm is used for optimizing the number of buckets and the split values of the data. Once the data is processed into buckets, it can be fed into the decision tree. The results indicate that while both methods can be equally strong, the number of buckets is an important factor for determining the optimum decision tree.

Future studies that aim misclassification error reduction may involve improved ANN techniques such as dynamic networks, etc. Hybridization of decision trees with other metaheuristics and changing the decision tree structure are also worth analyzing. Additionally, different churn data would be helpful for the validation of the results. This study does not involve temporal predictors of churn; these predictors should be integrated into the churn model.

REFERENCES

- [1] M. Xu and J. Walton, "Gaining customer knowledge through analytical CRM", *Industrial Management & Data Systems*, vol. 105, no. 7, pp.955 - 971, 2005.
- [2] A. Niedziółka, "Management of Agritourism in the Sustainable Development of Rural Areas with the example of the Maiopolskie Voivodeship", *The Journal of Management and Sustainable Development*, vol. 17, pp. 70-75, 2007.
- [3] C.P. Wei and I.T. Chiu, I.T., "Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach", *Expert Systems with Applications*, vol. 23, no. 2, pp. 103-112, 2002.
- [4] W. Verbeke, K. Dejaeger, D. Martens, J. Hur and B. Baesens, "New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach", *European Journal of Operational Research*, vol. 218, pp. 211–229., 2012.
- [5] J. Hadden, A. Tiwari, R. Roy and D. Ruta, "Churn Prediction: Does Technology Matter?", *International Journal of Intelligent Technology*, vol. 1, no.2, p. 104, 2006.
- [6] W.H. Au, K.C.C. Chan and X. Yao, "A Novel Evolutionary Data Mining Algorithm With Applications to Churn Prediction", *IEEE Transactions On Evolutionary Computation*, vol. 7, no. 6, pp. 532-545, 2003.
- [7] I. Khan, I. Usman, T. Usman, G. Ur Rehman and A. Ur Rehman, "Intelligent Churn prediction for Telecommunication Industry", *International Journal of Innovation and Applied Studies*, vol. 4, no. 1, pp. 165-170, 2013.
- [8] Y. Xie, X. Li, E.W.T. Nigai and W. Ying, "Customer churn prediction using improved balanced random forests", *Expert Systems with Applications*, vol.36, no. 3, pp. 5445-5449, 2009.
- [9] A. Idris, M. Rizwan and A. Khan, "Churn Prediction in Telecom Using Random Forest and PSO Based Data Balancing in Combination with Various Feature Selection Strategies", *Computers & Electrical Engineering*, vol. 38, no. 6, pp. 1808-1819, 2012.
- [10] T. Vafeiadis, K.I., Diamantaras, G. Sarigiannis and K.Ch. Chatzivasvas, "A comparison of machine learning techniques for customer churn prediction", *Simulation Modelling Practice and Theory*, vol. 55, pp.1-9, 2014.
- [11] J.H. Ahn, S.P. Han and Y.S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry", *Telecommunications Policy*, vol. 30, pp. 552-568, 2006.
- [12] L. Bin, S. Peiji and L. Juan, "Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System Service", *International Conference on Service Systems and Service Management*, Chengdu, China, June 9-11, pp. 1-5, 2007.
- [13] B. Larivière and D. Van den Poel, "Predicting customer retention and profitability by using random forests and regression forests

- techniques", *Expert Systems with Applications*, vol. 29, no. 2, pp. 472-484, 2005.
- [14] A. Lemmens and C. Croux, "Bagging and boosting classification trees to predict churn", *Journal of Marketing Research*, vol. 43, no. 2, pp. 276-286, 2007.
- [15] W. Buckinx and D. Van den Poel, "Customer base analysis: partial defection of behaviorally-loyal clients in a non-contractual FMCG retail setting", *European Journal of Operational Research*, vol. 164, pp. 252-268, 2005.
- [16] K. Coussement and D. Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques", *Expert Systems with Applications*, vol. 34, no. 1, pp. 313-327, 2008.
- [17] C. F. Tsai and Y. H. Liu, "Customer churn prediction by hybrid neural networks", *Expert Systems with Applications*, vol. 36, no. 10, pp. 12547-12553, 2009.
- [18] A. Karahoca and D. Karahoca, "GSM Churn Management by Using Fuzzy C-Means Clustering and Adaptive Neuro Fuzzy Inference Skkkkystem", *Expert Systems with Applications*, vol. 38, no. 3, pp. 1814-1822, 2011.
- [19] G. E. Xia and W. D. Jin, "Model of Customer Churn Prediction on Support Vector", *Machine. Systems Engineering – Theory and Practice*, vol. 28, no.1, pp. 71-77, 2008.
- [20] J. R. Quinlan, "Induction of decision trees", *Machine Learning*, vol. 1, no. 1, pp. 81-106, 1986.
- [21] K. Gajowniczek, T. Zabkowski and A. Orłowski, "Comparison of Decision Trees with Renyi and Tsallis Entropy Applied for Imbalanced Churn Dataset", *Proceedings of the Conference on Federated Computer Science and Information Systems*, 13-16 Sept., Lodz, pp.39-44, 2015.
- [22] J. Du, R. He and Z. Zhechev, "Forecasting Bike Rental Demand", *CS 229 Machine Learning Project*, Stanford University, 2014.
- [23] J. Kennedy and R. Eberhart, "Particle Swarm Optimization", *IEEE International Conference on Neural Networks*, vol. 4, pp. 1942-1948, Perth, WA, 27 November-01 December, 1995.
- [24] S. Haykin, *Neural Networks and Learning Machines*, 3rd Edition, Prentice Hall, New Jersey, USA, 2008.
- [25] K. Pytel, T. Nawarycz, W. Drygas, "Anthropometric Predictors of Artificial Neural Networks in the Diagnosis of Hypertension", *Proceedings of the Conference on Federated Computer Science and Information Systems (FEDCSIS)*, 13-16 Sept., Lodz, pp. 287-290, 2015.