

PAPER • OPEN ACCESS

## Comparison of ANN Classifier to the Neuro-Fuzzy System for Collusion Detection in the Tender Procedures of Road Construction Sector

To cite this article: Hubert Anysz *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **471** 112064

View the [article online](#) for updates and enhancements.

# Comparison of ANN Classifier to the Neuro-Fuzzy System for Collusion Detection in the Tender Procedures of Road Construction Sector

Hubert Anysz<sup>1</sup>, Andrzej Foremny<sup>1</sup>, Janusz Kulejewski<sup>1</sup>

<sup>1</sup> Warsaw University of Technology, Department of Civil Engineering,  
ul. L. Kaczyńskiego 16, 00-637 Warsaw, Poland

h.anysz@il.pw.edu.pl

**Abstract.** As the contracts in the road construction sector in Poland are usually of extremely high value and financed from the state budget, the tender procedures should not allow for the non-concurrent behaviours of tender participants. Otherwise, the clients' losses will be of high value too. A database comprising hundreds of bidding procedures in the road construction industry in Poland has been developed. It includes the tender's participants, the locations of the roads sections, bids' values, the winners, and the types of roads. Every procedure has been evaluated and assigned to the set with a given level of collusion occurrence probability. The evaluation has required the analysis and the transformation of – described in the literature – collusive types of behaviours to the parameters of procedures that can be shown as numbers or ranks. Four criteria of a collusion threatened contracts have been chosen and applied for evaluation. Then, two methods of machine learning were applied. The first method was to train an artificial neural network (ANN) to classify the procedures to the aforementioned sets. The other method was to utilize artificial neural networks predictive capabilities enriched by the fuzzy sets theory. The multiple output from ANN was defined as membership function values. The use of the fuzzy sets theory – the process of defuzzification – helps to classify the tender procedures to the sets of different level of risk (of collusion appearance). The results achieved in these two separate processes are compared and discussed. The created tool can be applied for the future tender procedures as a pre-test of a collusion appearance.

## 1. Introduction

Collusion takes place when market prices are close to monopoly prices despite market structure is an oligopolistic one. Anti-competitive agreements are not a new phenomenon in the world economy. The idea of agreeing prices by producers is simple and was present in the past.

Authors have performed a thorough literature study and have not found complex self-learning analytical tool for collusion detection in public tenders for roads. Nowadays, artificial intelligence can be a useful tool in construction industry management for: company strategy prognosis [1], bidding decision making [2], [3], construction delay prognosis [4–6] and cost calculation [7]. Artificial intelligence can also be used for collusion detection [8] and harm quantifying [9]. However a significant number of researches was conducted concerning collusion detection using traditional methods [10–14].



Road tenders, as large and valuable construction contracts, are prone to collusion [15]–[18]. Social and economic effects of tender collusion in road construction are particularly dangerous for society due to the enormous social losses that they can cause.

Authors in this research tested and compared ANN classifier and the hybrid neuro-fuzzy system as tools for public tender's collusion detection in road sector.

## **2. Input data description**

At first, a list of 569 public tenders was prepared using polish Public Procurement Bulletin (BZP) and/or European journal TED (Tenders Electronic Daily). Authors selected only polish tenders for road construction or reconstruction that had results published between 01.07.2014 and 30.06.2017 and had the winning tender offer value 10 million PLN or more (including VAT). The idea was to select the most valuable and important tenders where the biggest general contractors take part. As a next step, a list of 320 tenders was eliminated from the database because their scope was not in the area of authors' interest (road works were insignificant part of the whole order), or they missed cost/other necessary data, that according to the law, should have been published on the ordering party website but they were not.

The final database in this study consists of 249 tender records that all cover below mentioned data: participants' names, the locations of the roads sections, bids' values, order formula (build/design and build), scope of work (construction/reconstruction), type of road (highway/express/voivodeship or county).

## **3. Tenders evaluation and attribution of collusion possibility**

Authors likewise in the previous study [8] basing on official publications of OECD (Organisation for Economic Co-operation and Development) and UOKiK (The Office of Competition and Consumer Protection) selected and applied four criteria of searching for collusion behaviours in the prepared database [16], [18], [19] : the number of offers placed in a tender, the range of bid prices being offered, repeating ordering the works execution to the same contractor in a given location, the same set of offers in a few tenders. Two of these criteria were previously used [9] for estimating potential losses of the ordering party in public procurement in case of collusion.

Each of 249 procedures from the database has been evaluated and assigned to the set with a given level of collusion possibility. The evaluation has required the analysis and the transformation of – described in the literature – collusive types of behaviours to the parameters of procedures that can be shown as numbers or ranks. Four criteria of a collusion threatened contracts have been chosen and applied for evaluation.

### **3.1. Criterion 1. The number of offers placed**

A relatively small number of offers in a single tender can be a sign of collusion. In the analysed database the most frequent number of the offers placed in a single tender was 5 (in 45 tenders). The minimal number of offers was 2, the maximal number of offers was 20, the average number of offers was 6,9. Authors selected 51 tenders that had 4 or less offers placed in each as fulfilling criterion no. 1 of collusion possibility.

### **3.2. Criterion 2. The range of bid prices**

A small range of bid prices in comparison to other, similar tender procedures may signalize higher possibility of collusion. The range  $R_i$ , was calculated using the same formula likewise in the previous research [10]. In the analyzed database the average range of bid prices was 13,5 %, maximum range was 35,9 %. Authors selected 34 tenders that had range of bids equal to 5 % or less as fulfilling criterion no. 2 of collusion possibility.

### **3.3. Criterion 3. Geographical**

Three companies (called together as BIG3) has won 48,2 % of all analysed tender procedures. Extending the number to seven companies (BIG7) it can be said, that they had 61,0 % of the road construction market in Poland. The data concerning the numbers of procedures won are shown in table 1.

**Table 1.** Repetitive success in tender procedures

Number of procedures won	1	2 to 5	6 to 10	37 to 45
<b>Number of companies</b>	44	19	4	3

It was assumed, that non-concurrent tenders can appear only if a given company is able to win a few times (not only one or two times). Then the competitors recognize it as a strong player on the market and they can try to protect their business by illegal practises. Table 2 shows Polish voivodships where the market share of BIG3 or BIG7 was far from the average (calculated for the whole country; 48,2 % and 61,0 % respectively).

**Table 2.** Irregularities in BIG3 or BIG7 market share by voivodships

Voivodship by its capital	Market share of BIG3	Market share of BIG7	Total number of tenders
<b>Białystok</b>	8.3 %	8.3 %	12
<b>Bydgoszcz</b>	20.0 %	50.0 %	10
<b>Zielona Góra</b>	80.0 %	80.0 %	5
<b>Lublin</b>	83.3 %	83.3 %	6
<b>Katowice</b>	15.8 %	26.3 %	19
<b>Warszawa</b>	63.0 %	80.2 %	81

All procedures from Białystok and Katowice were assumed as fulfilling the criterion due to extremely low market share of BIG3 or BIG7. In Bydgoszcz, only the procedures where one of BIG3 has won were taken into consideration. All procedures where the winner was from BIG7 were taken into account in Zielona Góra, Lublin and Warszawa.

### 3.4. Criterion 4. The coincidence of the set of participants and the winners

The association analysis (a module built in Statistica 13 software) is one of the best tool for searching coincidence in a big set of data or when number of cases is not so big, but every single case is described by many features. Following 16 features were selected to describe every tender procedure:

- the average value of bid prices (0 for lower than 100 million PLN, 1 for higher value),
- the number of tender participants (0 for 5 or less, 1 for more),
- the exclusion of any bid from tender (0 for none, 1 if any offer was excluded),
- the winner from BIG3 (1 if any from the companies called together as BIG3 has won, 0 for any other winner),
- the winner from BIG7 (1 if any from the companies called together as BIG7 has won, 0 for any other winner),
- the participation of each company from BIG7 (1 for placing an offer, 0 for not); it has created 7 data types,

- the order formula (0 for build, 1 for design and build),
- the type of road (4 types were labelled),
- the scope of works (0 for reconstruction, 1 for building a new road),
- voivodship (16 voivodships were labelled).

The rule finding in association analysis means finding *bodies* ( $b$ ) and *heads* ( $h$ ) [20] maximizing the value of proportions called support ( $sup$ ) and confidence ( $conf$ ) [21].

$$sup(b \rightarrow h) = \frac{n(b \cap h)}{N} \quad (1)$$

$$conf(b \rightarrow h) = \frac{n(b \cap h)}{n(b)} \quad (2)$$

where:

$n(b \cap h)$  the number of cases when  $b$  and  $h$  appear simultaneously,  
 $N$  the total number of examined cases,  
 $n(b)$  the number of cases when  $b$  appears (regardless the head).

The phenomena or parameters of every procedure – *bodies* and *heads* – can be analysed as singular or they can be grouped. Proceeding the calculations number of *bodies* was limited to 10 and the *head* was limited to 1 feature. Despite setting the minimum  $conf$  as 70 %, over 2500 rules were found by the software. There were several rules showing the strong coincidence of participation in the tender procedures of the companies from BIG3 (let's label them C1, C2 and C3). The exemplary rules are shown in table 3.

**Table 3.** The exemplary rules found

The rule number	body	then head	conf %	sup %
1	C2=0, C3=0	→ BIG3=0	76,2 %	14,1
2	C2=1, C3=1, C5=0, C6=0	→ C1=1	100,0 %	10,8
3	C1=1, C2=1, C3=1, value=0, and type=V	→ BIG3=1	73,5 %	10,0
4	C1=1, C3=1, type=V	→ BIG3=1	70,8 %	13,7
5	C1=2, C3=1, type=V	→ BIG3=1	72,2 %	10,4
6	C2=1, C3=1, type=V, value=0	→ C1=1	100,0 %	13,7

The rule number 1 can be read as follows: if the company C2 has not placed an offer, so the company C3, then the head has happened i.e. none of the companies creating BIG3 has won. This coincidence has been found in 14,1 % of procedures. In 76,2 % of cases where C2 and C3 have not placed offers, C1 (the third from BIG3 group) hasn't won too. The rule number 6 is very strong. Every time ( $conf. = 100\%$ ) C2 and C have placed their offers for voivodship type of road and the average value of offers was below 100 million PLN, C1 has placed the offer too. It had a place in 13,7 % of procedures. The rules shown in table 3 as well as several other found, were the base of assumption, that criterion number 4 is fulfilled when the company C1, C2 and C3 has placed their offers for voivodship type of road.

### 3.5. The evaluation of the possibility of collusion occurrence

Having all four criteria set, the whole database of 249 procedures have been checked, which of tenders fulfil which collusion criteria. The result is shown in table 4. The next step was to determine how many criteria has been fulfilled by every tender from the collected database. The result is shown in table 5.

**Table 4.** Number of tenders that fulfilled given criterion

Criterion number	Criterion No.1	Criterion No.2	Criterion No.3	Criterion No.4
<b>Number of tenders</b>	66	34	108	33

**Table 5.** Number of tenders that fulfilled given number of criteria

Number of criteria fulfilled	0	1	2	3	4
<b>Number of procedures</b>	78	111	51	8	1

Having that determined, it was decided to divide the procedures onto three subsets. Procedures that have fulfilled only one criterion or none of criteria were assigned to the subset “Free from collusion”. As it was written in [9] fulfilling 2 criteria can create serious suspicion that non-concurrent behaviours of tender participants has occurred (subset “Collusion suspected”). Then 3 or 4 criteria are fulfilled, the tender was assigned to the subset “Collusion highly expected”.

**Table 6.** Classification

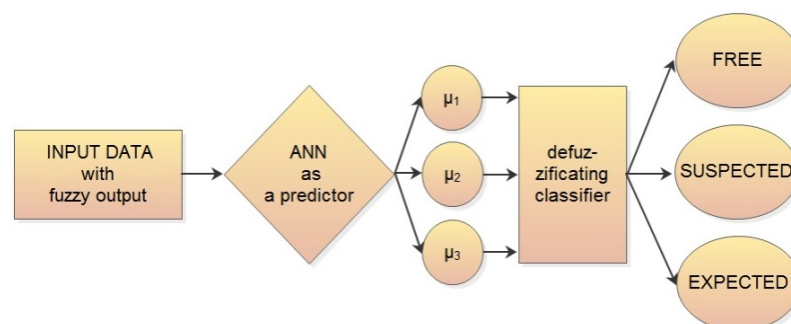
Subsets	Free from collusion	Collusion suspected	Collusion highly expected
<b>Number of criteria fulfilled in a given procedure</b>	0 or 1	2	3 or 4

Described above set of processes of evaluating tender procedures has allowed for classification of every procedure to one of the three aforementioned subsets. They are the reference base used in the next step for automatic classification of tender procedures. It is to emphasise, that the classification shown in the table 6 it is only the authors’ opinion. The only institution entitled to state if a collusion really had a place is a court.

#### 4. The automatic detection of collusion

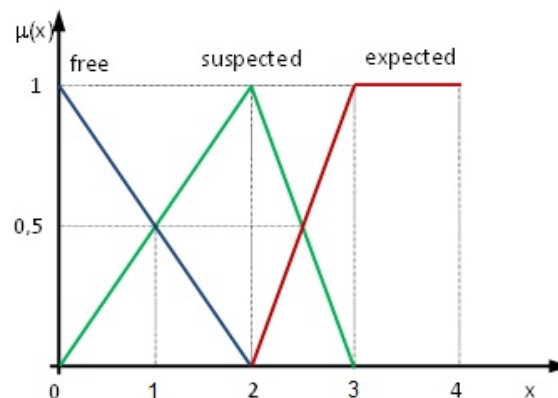
The following two methods of automatic classification have been applied:

- neuro-fuzzy system (see figure 1) that utilizes predictive features of ANN,
- classification features of ANN.

**Figure 1.** Neuro-fuzzy classifying system

#### 4.1. The neuro-fuzzy system

Despite the fact the original classification was made according to authors' opinion, it has to be taken into consideration that the tender procedure fulfilling 2 criteria can be free from any non-concurrent processes. Another procedure fulfilling only 2 criteria can be illegal according to collusion made by its participants. That is why the fuzzyfication was made (according to figure 2). For each procedure the value of three membership  $\mu(x)$  function was determined. They describe the level of membership to the three fuzzy subsets called "free from collusion", "collusion suspected", "collusion highly expected". The domain of  $\mu(x)$  was determined based on table 5 and shown in figure 2.



**Figure 2.** The membership functions of number of criteria fulfilled

Then the ANN was found for simultaneous predicting all three membership values. The input data for ANN were similar to the data for association analysis, but average value of the offers, their range and the number of participants in each tender procedure were taken as a standardized numbers. The linear-maximum method for standardization was used. According to the limited number of cases in the subset "collusion highly expected" the data was divided manually into teaching, validating and testing, to provide the representation of each output subset in each type of input data. The following ANN (multilayer perception type) 16-16-3 with logistic activation function in a hidden layer and exponential activation function in an output layer has produced the lowest MSE (mean squared error). The predicted values of membership functions, for validating and testing data (70 cases; 28 % of all cases) were defuzzificated with the use of following three methods: centre of gravity, weighted average, maximum described in [22]. The result of classification (the best in this case, by the weighted average method) is shown in table 6.

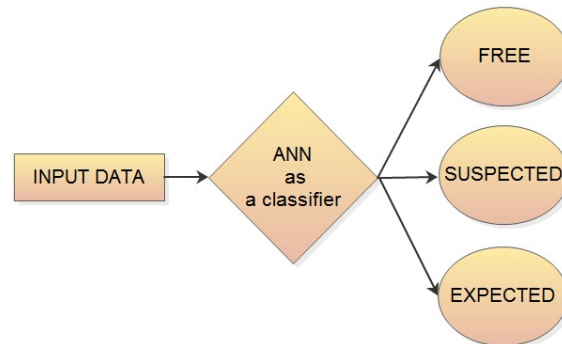
**Table 7.** Classification results of neuro-fuzzy system

subsets	Free from collusion	Collusion suspected	Collusion highly expected	All subsets
<b>All cases</b>	52	14	4	70
<b>Correct</b>	34	10	1	45
<b>Incorrect</b>	18	4	3	25
<b>Correct [%]</b>	65,38	71,43	25,00	64,28
<b>Incorrect [%]</b>	34,62	28,57	75,00	35,72

#### 4.2. Tenders classification using ANN

The second method was to train an artificial neural network to classify the tenders to the aforementioned sets. Calculations were made using software Statistica 13, minimal number of neurons

in the hidden layer was 3, maximum was 30. The same four activation functions were used in the hidden and output neurons: linear, logistic, hyperbolic tangent and exponential.



**Figure 3.** ANN as a classifier

One network was found that classified 100 % of records correctly but only in the subset “free from collusion”, where in classes “collusion suspected” and “collusion highly expected” it was only 50 % correctly found each. The best classifying network had quite high average accuracy 90 %. It classified correctly a significant number of 96 % subset “free from collusion” records, classes “collusion suspected” and “collusion highly expected” results were above 70 % each. This network had 16 neurons in the input layer, 12 neurons in the hidden layer and 3 neurons in the output layer. Activation function in the hidden layer was hyperbolic tangent and activation function in the output layer was normalized exponential function. Results of this network was presented in table 7 below.

**Table 8.** Classification results summary for the best artificial neural network

subsets	Free form collusion	Collusion suspected	Collusion highly expected	All subsets
<b>All cases</b>	52	14	4	70
<b>Correct</b>	50	10	3	63
<b>Incorrect</b>	2	4	1	7
<b>Correct [%]</b>	96,15	71,43	75,00	90,00
<b>Incorrect [%]</b>	3,85	28,57	25,00	10,00

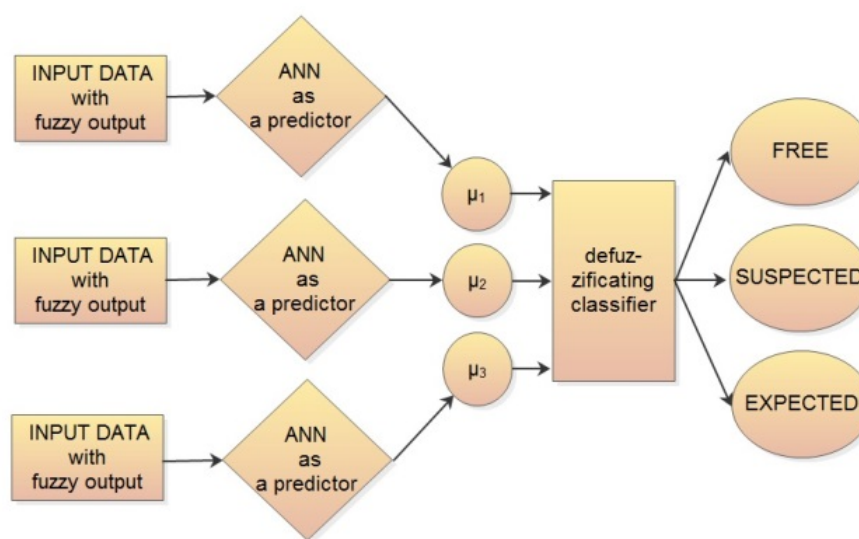
## 5. Results and discussions

The comparison of results achieved, summarized in table 6 and table 7 clearly shows, that neuro-fuzzy system applied for collusion detection (based on number of 4 collusion criteria fulfilled) is not so effective as pure ANN applied for classification. Predictions of membership functions made by ANN with fuzzy classifier has produced 64,3 % correct classifications (calculated for testing and validating datasets), while classifying ANN has given 90,0 %. Analysing 3 subsets separately, superiority of ANN classification over neuro-fuzzy system seems higher. Searching for collusion, the analyst can exclude the cases which are free from collusion, found with 96,15 % correctness (65,38 % in neuro-fuzzy classifier). After that, proving the collusion existence can be started with tender procedures comprised by the subset “collusion highly expected”, which are found with 75 % correctness (25,00 % in neuro-fuzzy classifier). The fact that both tools has attributed the tender procedures to the subset “collusion suspected” become not important for evaluation of applied tools. Much lower effectiveness of neuro-fuzzy classifier (compared to ANN classifier) can arise from varying much number of record



in database for each subset (189, 51 and 9 respectively). Dividing 9 cases into teaching, validating and testing subsets of data, makes machine recognizing really difficult. Taking that into consideration, performance of ANN classifier (90 % overall correctness and 75 % for 9 element subset) can be evaluated as very effective.

It has to be mentioned, that during the calculation and analysis proceeded, another classifying ANN has been found, that has attributed 100 % correctly to the subset “free from collusion”, but their overall performance was lower than ANN described in point 4.2. Another important (for future user of automatic collusion detection) result has been achieved. The authors have tried to modify neuro-fuzzy system, by predicting three membership functions with three separate ANN (with only one output neuron each; see figure 4).



**Figure 4.** Modified neuro-fuzzy classifying system

Due to the low overall correctness of classifying, the system has not been analyzed. The level of correctness of attributing to the subset “collusion highly expected” was low too (0,0 %). The interesting feature was, that analysing membership function  $\mu_3$  separately and setting its lower limit to 0,20, it has attributed correctly all (4) cases fulfilling 3 or 4 criteria (from validating and test data) to the subset “collusion highly expected”. At the same time, 3 other tender procedures were wrongly attributed to this set (so the correctness level in this subset was 57,1 %). Nevertheless, described above applying of modified neuro-fuzzy classifying system could help the analyst not to omit any procedure where collusion is highly expected.

## 6. Conclusions

It has to be remembered that existence of a collusion in a given procedure can be stated only in a court judgement. The database – with assigned levels of collusion possibility to the real tender’s procedures – it is only the authors’ opinion. Based on that, it was confirmed that with the use of artificial neural networks, the tender procedures can be effectively classified (90 % correctness) to the subsets described as “free from collusion”, “collusion suspected” and “collusion highly expected”. The neuro-fuzzy system has not produced so good results, but it can be used (or its elements) in refining results from ANN classifier. The low correctness of neuro-fuzzy classifier cannot be generalized. When the number of elements of output subsets is more balanced (unlike in analysed case), its performance can be much better.

## References

- [1] Kulejewski J., „Rozpoznawanie i prognozowanie strategii przetargowej przedsiębiorstwa budowlanego”, *Recognition and forecasting of a tender strategy of a construction company*, Proc. of Theoretical Foundation of Civil Engineering, 2006, s. 347–356.
- [2] Leśniak A., Wspomaganie decyzji wykonawcy budowlanego z zastosowaniem sztucznej inteligencji, *Supporting the decision of the construction contractor using artificial intelligence*, *Czasopismo inżynierii lądowej, środowiska i architektury*, t. 33, nr z.63 (1/I/16), s. 189–196, 2016.
- [3] Leśniak A., i Plebankiewicz E., *Modelling the Decision-Making Process Concerning Participation in Construction Bidding*, *Journal of Management in Engineering*, t. 31, nr 2, s. 4014032, 2015.
- [4] Anysz H., „Wykorzystanie sztucznych sieci neuronowych do oceny możliwości wystąpienia opóźnień w realizacji kontraktów budowlanych”, *The use of artificial neural networks to assess the possibility of delays in the implementation of construction contracts* Politechnika Warszawska, 2018.
- [5] Anysz H., Zbiciak A., i Ibadov N., *The Influence of Input Data Standardization Method on Prediction Accuracy of Artificial Neural Networks*, *Procedia Engineering*, t. 153, nr August, s. 66–70, 2016.
- [6] Rogalska M., *Wieloczynnikowe modele w prognozowaniu czasu procesów budowlanych*, *Multifactorial models in forecasting the time of building processes*, ISBN: 978-83-7947-186-7, March 2016.
- [7] Juszczak M. B., i Zima K., *ANN Based Approach for Estimation of Construction Costs of Sports Fields*, t. 2018, 2018.
- [8] Foremny A., i Anysz H., *The collusion detection in public procurements – selected methods applied for the road construction industry in Poland*, 2018; 10.13140/RG.2.2.16082.04807.
- [9] Anysz H., Foremny A., i Kulejewski J., *Estimating potential losses of the client in public procurement in case of collusion utilizing a MLP neural networks*, *Technical Transactions*, t. 111, nr 1–B, s. 105–118, 2014. 10.13140/2.1.2409.6321.
- [10] Porter R. H., i Zona J. D., *Detection of Bid Rigging in Procurement Auctions*, *Journal of Political Economy*, t. 101, nr 3, s. 518–538, 1993.
- [11] Kenneth H., i Porter R., *Collusion in auctions*, *Annales d’économie et de statistique*, t. 15, 1989.
- [12] Harrington J., *Behavioral screening and the detection of cartels*, *European Competition Law Annual*, 2006.
- [13] Zona J. D., *Bid-rigging and the Competitive Bidding Process: Theory and Evidence*, Ph.D. Dissertation, State University, 1986.
- [14] Gabrielli M. F., *Detecting Collusion on Highway Procurement*, *Económica*, t. LIX, January 2013, s. 1–31, 2013.
- [15] Oke A., Aigbavboa C., i Mangena Z., *Prevention of Collusion for Innovative Construction*, *Procedia Engineering*, t. 196, nr June, s. 491–497, sty. 2017.
- [16] Urząd Ochrony Konkurencji i Konsumentów, „Zmowy przetargowe”, *Tender collusion*, 2017.
- [17] Organisation for Economic Co-operation and Development (OECD), „Fighting bid rigging in public procurement: Report on implementing the OECD Recommendation”, 2016.
- [18] Organisation for Economic Co-operation and Development (OECD), *OECD recommendation of the council on fighting bid rigging in public procurement*, 2012.
- [19] Organisation for Economic Co-operation and Development (OECD), *Guidelines for fighting bid rigging in public procurement. Helping governments to obtain best value for money*, 2012.
- [20] T.Morzy, *Eksploracja danych. Metody i algorytmy*, *Data mining. Methods and algorithms* Wydawnictwo Naukowe PWN, 2013.
- [21] Larose D. T., i Larose C. D., *Discovering Knowledge in Data*. 2014.
- [22] Anysz H., Ibadov N., *Neuro-fuzzy predictions of construction site completion dates*, *Czasopismo Techniczne*, t. 6, s. 51–58, 2017; DOI 10.4467/2353737XCT.17.086.6562.