

# Using artificial intelligence to analyse businesses in agriculture industry

Jakub Horák<sup>1,\*</sup>

<sup>1</sup>University of Žilina, Faculty of Operation and Economics of Transport and Communications, Univerzitná 1, 01026 Žilina, Slovak Republic

**Abstract.** Artificial intelligence is largely used in many technical applications and allows you to provide various solutions in problem estimation, regression, or optimization. Artificial intelligence, specifically artificial neural networks, extend to the area of economics and finance. They are used primarily for operations that can't be identified analytically. Neural networks are suitable for modelling very complex strategic decisions, for large sets of data, and so on. The main advantage is the ability to learn and then to capture hidden and strongly non-linear dependencies. In this paper they are used for the analysis of agricultural businesses. The aim is to analyse the state of the agricultural sector through the use of Kohonen networks and then to assess its future development. On the basis of the analysis, significant and large clusters of businesses are depicted, and the most significant clusters are analysed. It is possible to estimate the number of businesses that will be successful, those that will stagnate and those that will fail in the following period. Application of Kohonen networks is rather complex, but they have great potential and the results are very interesting.

**Key words:** Kohonen networks, cluster analysis, prediction, agricultural businesses, sector development

## 1 Introduction

Artificial neural networks (ANNs) try to copy processes in the human brain and the nervous system through computer devices. The concept of an artificial neural network first appeared in biology and psychology. According to Klieštík [1], ANNs are computational models that are inspired by biological neural networks, namely neuronal behavior. The use of ANNs, however, crosses over to other areas as well. Currently, these networks are widely used to address possible future problems, to predict values [2]. Altun, Bilgil and Fidan [3] claim that ANNs are extensively used in many technical applications and provide various solutions in problem estimation, regression, or optimization. ANNs are used for various purposes and it is appropriate to use them for demanding operations that can't be identified analytically. Guresen and Kayakutl [4] state that they are therefore mainly used to model very complex strategic decisions. According to Sánchez and Melina [5], ANNs can be applied to function approximations, classification, and prediction of time series. They help us understand functions that exist in model groups and environments that are constantly changing. ANNs

---

\* Corresponding author: horak@mail.vstecb.cz

are able to offer solutions to complex models of improved artificial intelligence. Beiranvand et al. [6] maintain that the results of ANNs are very promising, and their performance and accuracy when solving a company's key performance indicators are considerably higher than those of traditional statistical techniques.

The advantage of ANNs is the ability to work with large data, accuracy of results, or simpler use of the acquired neural network [7]. ANNs are a group of intelligent data analysis technologies that differ from other classical techniques [8]. According to Santin [9], ANNs have many advantages over conventional methods. They are able to analyse complex patterns very quickly and with high precision and are flexible in their own use. Sayadi et al. [10] contend that some of the main advantages of neural network methods for predicting key company indicators are the ability to learn, the ability to generalize, etc. The disadvantage of ANNs is their demand for large sample data, because to create such an amount of data, many test trials are needed, which is not ideal for the user. Another disadvantage is the process of optimizing topology of hidden layers, which is time consuming, complicating the computation process [11]. Rowland and Vrbka [12] consider ANNs to be particularly disadvantageous in that they require high data quality and architecture definition, the possibility of illogical network behavior.

One of the ANN models is the Kohonen Network, which can be used to group sets of data into different groups. The data is grouped so that the records within the group tend to be similar to each other, and the records in the different groups are different. Many experimental results show that Kohonen's networks are very effective for assessing companies [13]. The Kohonen network consists of an input layer that is completely interconnected with the output layer and learns by self-organization. This network has a wide range of uses because it is an alternative network that is usable for most of the ANNs calculations. It is used primarily for speech processing, audio editing, security applications, photos, videos, and allows the projection of high-dimensional data into lower-dimension data [14]. Many experts even claim that Kohonen's maps are a highly versatile tool, and if one does not know what method to use, he can use the Kohonen network with a quiet conscience. There is no need for the author to know about the information he processes and the network will categorize it by itself [15].

The Kohonen networks are used in this paper to analyse agricultural businesses. Like every branch of the national economy, agriculture also has its own specificities. These consist mainly in satisfying basic human needs. Agriculture is a source of food of both animal and plant origin, which is necessary for human nutrition. It produces many raw materials for other industries. According to Waheed et al. [16], agriculture is primarily characterized by its maximum dependence on natural conditions, weather. Agricultural land is therefore the main production means for agricultural production. Agricultural production systems use sophisticated techniques to correlate human, natural, industrial and economic resources. This is done to meet the demand for food in today's highly competitive and demanding market in terms of environmental and social sustainability [17].

The aim of this article is to carry out a cluster analysis of enterprises operating in the area of agriculture.

## **2 Data and methods**

A data set of 4213 enterprises having operated in the area of agriculture in 2016 is made up for purposes of this article. The data set of an enterprise is generated from Albertina database. Comprehensive financial statements of these companies (all financial data are provided in thousands CZK) are available. The analysis deals only with some of the items:

1. Total assets: the size of the enterprise, volume of assets (fixed, current and other assets) in particular

2. Fixed assets: this item refers to the degree of the participation of fixed assets in an enterprise and informs about the technology used in the production. The higher the participation of fixed assets is, the higher the supposed automated production should be.
3. Current assets: this item refers to the amount of financial resources, claims and inventories of the enterprise. They mostly concern working capital, i.e. assets that change their nature within a period shorter than one year and that immediately participate in main activities of the enterprise.
4. Equity: they refer to the degree of possible business risk that owners of enterprises take; they mostly concern the amount of equity that may be lost in case of insolvency.
5. Borrowed capital: they refer to the capital that is not associated with the participation in the commercial management of the company. They provide an external view of a potential success of the enterprise.
6. Performance: they refer to main activities of the enterprise – production.
7. Added value: it refers to a value added to the capital investment generated by the enterprise through its main activity.
8. Operating profit: it informs about the success rate of the enterprise in its main activity.
9. ROE: income of equity means the evaluation of the equity of the enterprise, i.e. the capital provided to the entrepreneur.
10. Economic result before taxation: it refers to a success or failure of the enterprise in its entire activity.

In this way, these items rather refer to characteristics, not value generators, which may have, though indirect, influence on the enterprise efficiency.

The data set is subsequently subjected to a cluster analysis, Kohonen networks in particular. The cluster analysis uses Statistica software Dell, version 12. Module Data Mining is applied as a tool for neural networks. Here, neural networks without teacher forcing are chosen – Kohonen networks. The analysis data are amassed; they all concern continuous predictors. The data set is divided into three parts:

1. Training data set: it involves 70% of enterprises from the data set. This data set is demonstrated by Kohonen network.
2. Testing data set: it involves 15% of enterprises from the data set. This data set helps verify parameters achieved from Kohonen network.
3. Validation data set: it involves 15% of enterprises from the data set. This data set also helps verify achieved Kohonen network – its applicability in particular.

The topological length and topological width of Kohonen map is determined to 10. The number of repetitions – iterations – is determined to 10,000. All the same, we should take into account that the error level is the most decisive factor. Unless the parameters of Kohonen network improve with individual iterations, the training will be terminated before 10,000th iteration. In case of a tangible improvement of parameters in 10,000th iteration, the whole procedure has to be repeated and a higher value of required iterations has to be generated in order to achieve the best possible result.

### 3 Results

According to the methodological part of the article, the data were divided into three data sets – training, testing and validation. Defining characteristics of all three data sets and the original data set are suggested in Tab. 1.

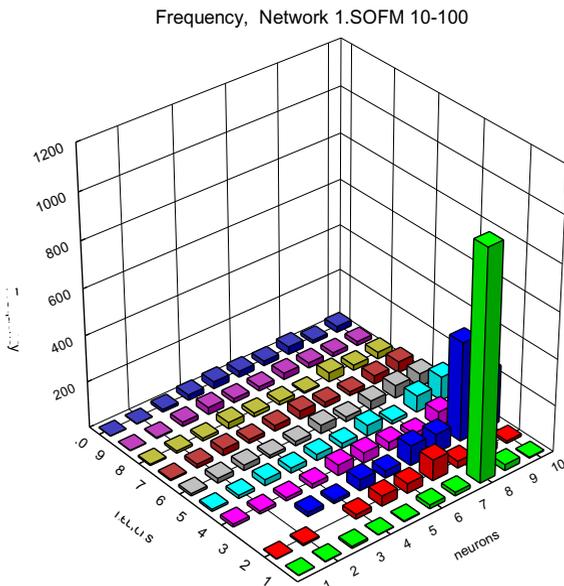
**Table 1.** Descriptive characteristics of the dataset

Samples	Total assets	Fixed assets	Current assets	Equity	Borrowed capital	Perform.	Added value	Operating profit	ROE	Economic result before taxation
Minimum (Training)	0	-325	-731	-517381	-3629	-3744	-21623	-33385	-4764.50	-27561
Maximum (Training)	703378	62152700	8122509	67905791	2295046	11844244	7831937	6356727	14086.50	6493751
Average (Training)	78961	54510	23916	55566	23176	31617	9518	4933	10.87	4709
Standard deviation (Training)	1302210	1148973	156285	1255363	68527	233653	145402	117554	321.28	120263
Minimum (Testing)	-2009	-1772	-2009	-24105	-8980	0	-9606	-19093	-1421.78	-44599
Maximum (Testing)	153017	539774	1117112	1274713	311733	1310205	384836	196452	1685.50	130947
Average (Testing)	54844	33103	21190	33664	20995	26175	6805	2744	4.64	3129
Standard deviation (Testing)	108531	62538	55404	82012	37082	68332	20271	9836	105.17	9642
Minimum (Valid.)	-460	-77	-3071	-76429	-623	-446	-12964	-27325	-3862.00	-31939
Maximum (Valid.)	896402	6843702	2105131	7744524	1209418	2248884	1439845	355045	2600.00	115722
Average (Validation)	72759	47430	24695	48527	23985	30873	9580	3103	0.03	2295
Standard deviation (Validation)	130054	85382	51325	102234	41875	66905	18555	10115	321.85	11109
Minimum (Overall)	-2009	-1772	-3071	-517381	-8980	-3744	-21623	-33385	-4764.50	-44599
Maximum (Overall)	703378	62152700	8122509	67905791	2295046	11844244	7831937	6356727	14086.50	6493751
Average (Overall)	74414	50238	23624	51225	22970	30689	9120	4330	8.31	4110
Standard deviation (Overall)	1099644	967693	137071	1057887	63981	201369	124025	98614	281.92	100738

Source: Author.

Ideally, the characteristics of individual data sets would be approximately identical. Nevertheless, regarding the fact that enterprises are randomly divided, that is not the situation. All the same, the random division of data does not necessarily have to have a negative impact on the result.

The calculation of Kohonen network was then carried out. The network is referred to as SOFM 10-100 further in the text. Enterprises were divided into separate clusters in a topological grid 10 x 10. The frequency of separate clusters is apparent from the three-dimensional diagram in Figure 1. The picture suggests that the highest number of enterprises in individual squares of the topological grid is situated in (1, 8); then squares (3, 9) follow and then (3, 10), (4, 10) and (5, 10). On the whole we might say that the highest number of 5,000 analysed enterprises is situated in sector (1, 7), (5, 7) and (1, 10), (5, 10).



**Fig. 1.** Diagram of the frequency of enterprises in the topological grid

Source: Author.

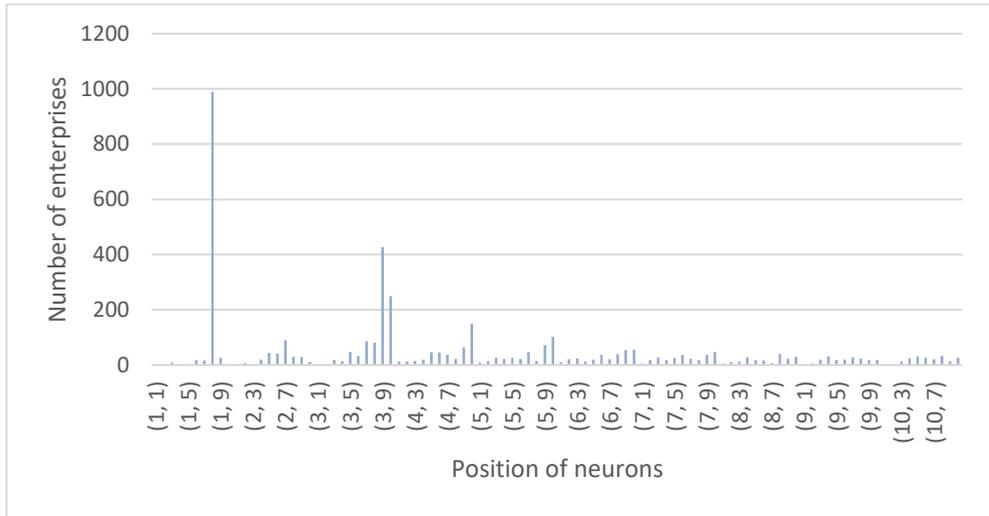
For more detailed results see Table 2.

**Table 2.** Frequencies of network SOFM 10-100

	Network: 1.SOFM 10-100 / Samples: Training, Testing, Validation									
	1	2	3	4	5	6	7	8	9	10
1	1	3	9	2	4	17	16	990	26	4
2	2	7	0	20	44	42	90	30	29	11
3	0	0	18	14	48	32	86	82	427	249
4	13	13	15	19	46	45	37	22	64	150
5	9	14	26	22	27	22	47	15	72	103
6	10	21	24	13	19	37	21	39	55	56
7	6	17	28	17	25	36	23	17	37	48
8	6	10	12	28	17	16	7	40	23	30
9	2	5	19	31	18	19	28	23	18	18
10	1	2	12	24	31	26	21	34	14	26

Source: Author.

The table suggests that our attention should be drawn to two clusters – (1, 8) and (3, 9) in particular. In regard to the calculating operations, these are very remarkable data sets, which represent more than one third of all enterprises in the data set; the dominance is therefore possible and may be evident from Kohonen diagram in Figure 2.



**Fig. 2.** Kohonen diagram for selected network cells SOFM 10-100

Source: Author.

Although the data set is relatively heterogeneous, the cluster analysis suggests certain similarities in some characteristics of the enterprises in question.

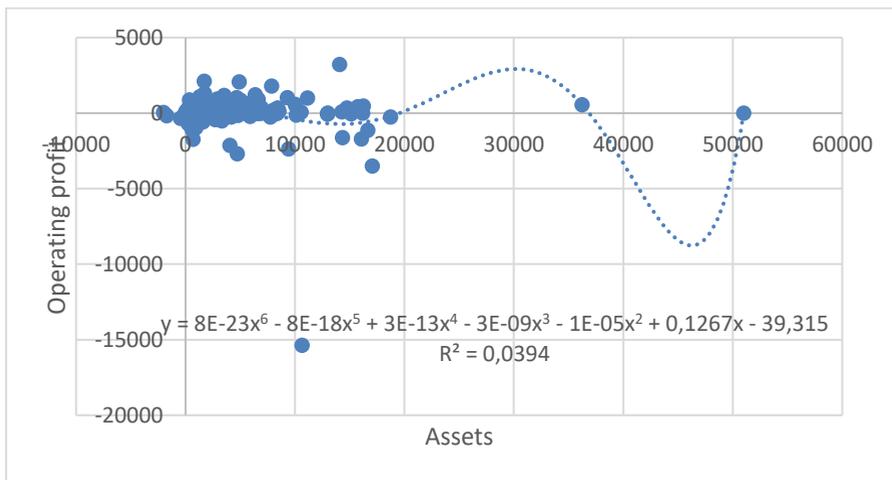
### **Cluster (1, 8)**

Cluster (1, 8) of the topological grid contains 990 enterprises. Its individual clusters constitute the biggest data set; the following average values might be derived from the input parameters:

1. Total assets: 1,284.2 mil. CZK,
2. Fixed assets: 687.8 mil. CZK,
3. Current assets: 583.1 mil. CZK,
4. Equity: 1,052 mil. CZK,
5. Borrowed capital: 225.2 mil. CZK,
6. Performances: 516 mil. CZK,
7. Added value: 138.2 tis. CZK,
8. Operating profit: 11.1 mil. CZK,
9. ROE: 27.6%.
10. Economic result before taxation: 290 mil. CZK.

As a matter of fact, a typical enterprise can be derived from average values of this cluster of enterprises. A small enterprise is here to be dealt with, whose operating profit is minimal. Despite this fact, its existence is not jeopardized. The degree of enterprise automation is relatively low. The enterprise holds a large amount of current assets (claims and inventories).

Provided a homogeneous cluster of enterprises is to be dealt with, the validity of its basic logical link needs to be assessed. Of interest may also be a close link between total assets and the amount of the operating profit (see Figure 3 for more details).



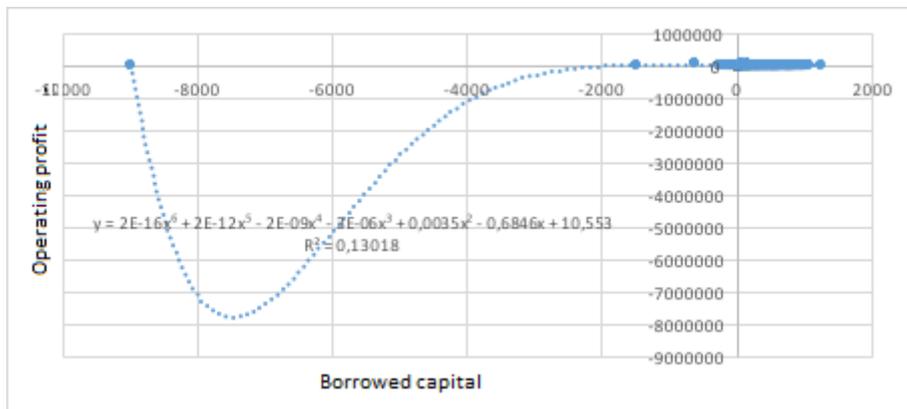
**Fig. 3.** Link of total assets and the operating profit of the cluster (1, 8)

Note: All values in graphs are in thousands of CZK.

Source: Author.

Total assets and operating profit correlate at more than 0.2; any significant correlation of both quantities is therefore ruled out. However, the data are provided with a regression curve – 6th degree polynomial in particular. Based on this polynomial the development of the operating profit can be assessed considering the size of the enterprise determined by the amount of total assets.

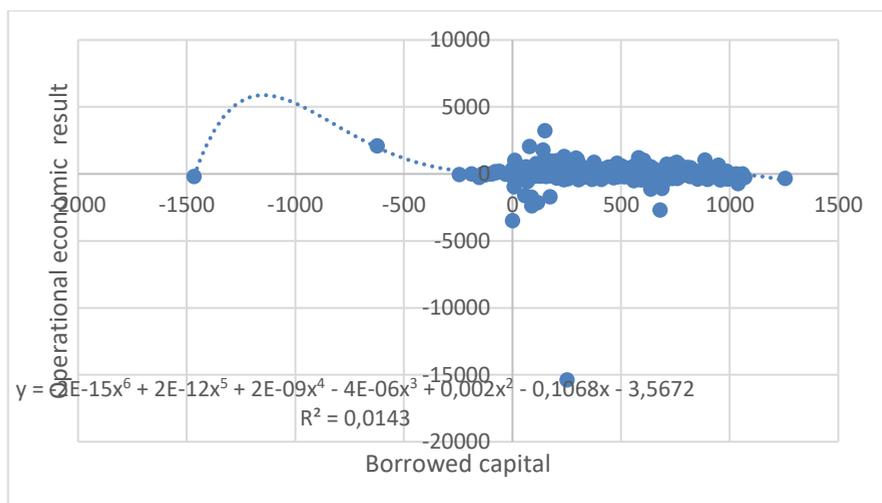
Figure 4 suggests a relationship between borrowed capital and operating profit.



**Fig. 4.** Borrowed capital and operating profit of cluster (1, 8)

Source: Author.

The figure shows that the data suggest an extreme value of a negative foreign capital and a very negative operational result at the same time; this extreme needs to be removed from the source data used in the diagram for further assessments of the entire data set. Zoo Tábor, plc. is here to be dealt with. Figure 5 illustrates the relationship between borrowed capital and operating profit without the extreme value.



**Fig. 5.** Borrowed capital and operating profits of cluster (1, 8) without the identified extreme value

Source: Author.

The relationship can show us financial leveraging in case of a relatively high correlation between both quantities; however, it amounts only to 0.36. Despite this fact, the diagram items are provided with 6th degree polynomial. This polynomial suggests that the optimal degree of enterprise debts situated in cluster (1, 8) is up to 1 mill. CZK.

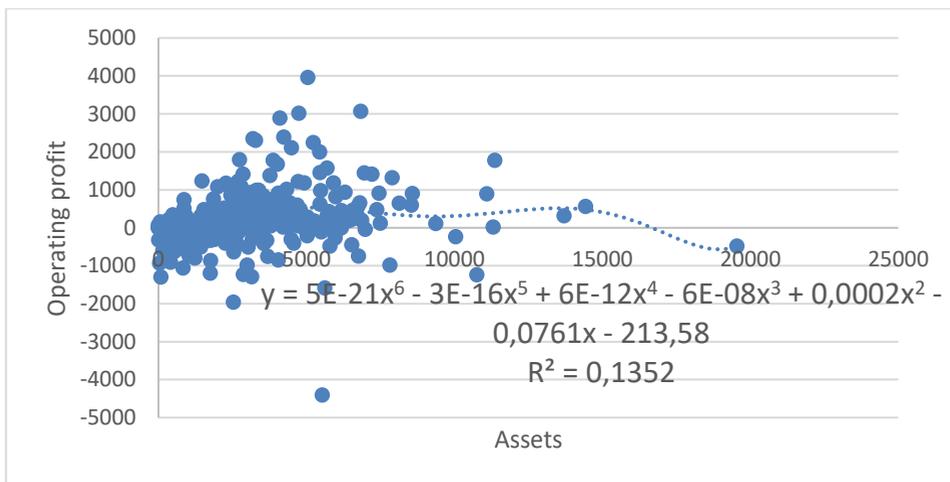
### Cluster (3, 9)

Cluster (3, 9) of the topological grid is ranked second as far as the frequency is concerned. It comprises 427 enterprises. In a view of input parameters, typical values of this cluster are as follows:

1. Total assets: 1,244.6 mil. CZK,
2. Fixed assets: 497.4 mil. CZK,
3. Current assets: 728.4 mil. CZK,
4. Equity: 518 mil. CZK,
5. Borrowed capital: 707.2 mil. CZK,
6. Performances: 1,174.8 mil. CZK,
7. Added value: 391 mil. CZK,
8. Operating profit: 85.4 mil. CZK,
9. ROE: 20.82%.
10. Economic result before taxation: 107.9 mil. CZK.

The typical enterprise of this cluster is twice as large as cluster (1, 8) (in a view of the amount of total assets). This enterprise is slightly less automated, but it possesses a higher amount of current and total assets. The biggest difference lies in the amount of the operating profit. The enterprise of cluster (3, 9) generates almost by 7% lower income of the equity – 20 in particular which means 82%. All the same, even this case is worth analysing of some of its variables.

Figure 6 suggests the relationship between total assets and the operating profit.

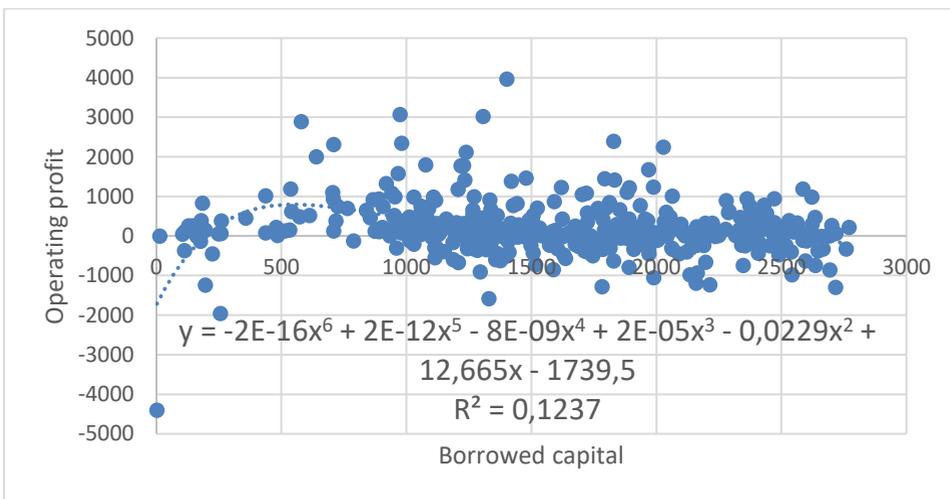


**Fig. 6.** Relationship between total assets and the operating profit of cluster (3, 9)

Source: Author.

The correlation coefficient of both quantities amounts to 0.38, which doesn't show any correlation between the two variables. However, this cluster manifests that a higher amount of assets means a better operating profit; the connecting line has again a shape of 6th degree polynomial.

Figure 7 depicts a relationship between borrowed capital and operating profit.



**Fig. 7.** Borrowed capital and operating profit of cluster (3, 9)

Source: Author.

Neither this case shows a high correlation coefficient. Nevertheless, even this example demonstrates that higher debts do not mean a better operating profit; on the whole, the total debt should not exceed 2 mil. CZK.

## 4 Conclusion

The aim of this article was to carry out a cluster analysis of enterprises operating in the area of agriculture. The specified aim was achieved only partially.

A cluster analysis using neural networks without teacher forcing – Kohonen networks – was carried out. According to the cluster analysis, enterprises were divided into clusters in Kohonen map (10 x 10 clusters). In regard to the number of enterprises, some of the clusters are of a major significance. Clusters (1, 8) and (3, 9) contain hugest numbers of enterprises. These clusters were subjected to a further analysis. Both clusters were provided with typical examples of these clusters. They differed in their size and the amount-generated operating profit. Enterprises of cluster (1, 8) were on the verge of profitability in 2016. The average operating profit for an enterprise amounted to 10 thousand CZK, which means minimal debts in the ideal world. Of interest might also be its degree of automated production. A typical enterprise of cluster (3, 9) is more than twice larger than that of cluster (1, 8). The participation of fixed assets was lower than in that of cluster (1, 8); this enterprise, on the contrary, holds a larger number of current assets. The average ROE of this enterprise was 20.82%. On the whole, the analysis pointed to two reasonable conclusions:

1. Larger enterprises (i.e. an enterprise with a larger number of assets) generate a better operating profit on average.
2. Financial leverage has a more positive effect in larger enterprises; i.e. the bigger the amount of borrowed capital is, the higher the operating profit and the income of equity are.

## References

1. T. Klieštk, Models of autoregression conditional heteroskedasticity garch and arch as a tool for modeling the volatility of financial time series. *Ekonomicko-manažerské spektrum*, 7(1), 2-10, (2013)
2. H. Pao, A comparison of neural network and multiple regression analysis in modeling capital structure. *Expert Systems with Applications*, 35(3), 720-727, (2008)
3. H. Altun, A. Bilgil, B. C. Fidan, Treatment of multi-dimensional data to enhance neural network estimators in regression problems. *Expert Systems with Applications*, 32(2), 599-605, (2007)
4. E. Guresen, G. Kayakutlu, Definition of artificial neural networks with comparison to other networks. *Procedia Computer Science*, 3, 426-433, (2011)
5. D. Sánchez, P. Melin, Modular Neural Networks for Time Series Prediction Using Type-1 Fuzzy Logic Integration. In: P. Melin, O. Castillo, J. Kacprzyk. *Design of Intelligent Systems based on Fuzzy Logic, Neural Networks and Nature-Inspired Optimization*, 141-145, (2015)
6. V. Beiranvand, A. Abu Bakar, Z. Othman, A Comparative Survey of Three AI Techniques (NN, PSO, and GA) in Financial Domain. *Proceedings of the 7th International Conference on Computing and Convergence Technology*, 332-337, (2012)
7. P. Šuleř, Using Kohonen's neural networks to identify the bankruptcy of enterprises: Case study based on construction companies in South Bohemian region. *Proceedings of the 5th International Conference Innovation Management, Entrepreneurship and Sustainability*, Prague, Czech Republic, pp 985-995, (2017)

8. A. Hogenboom, W. Ketter, J. Van Dalen, U. Kaymak, J. Collins, A. Gupta, Adaptive Tactical Pricing in Multi-Agent Supply Chain Markets Using Economic Regimes. *Decision Sciences*, **46**(4), 791-818, (2015)
9. D. Santin, On the approximation of production functions: a comparison of artificial neural networks frontiers and efficiency techniques. *Applied Economics Letters*, **15**(8), 597-600, (2008)
10. A. R. Sayadi, S. M. M. Tavassoli, M. Monjezi, M. Rezaei, Application of neural networks to predict net present value in mining projects. *Arabian Journal of Geosciences*, **7**(3), 1067-1072, (2012)
11. M. S. Hossain, Z. C. Ong, Z. Ismail, S. Noroozi, S. Y. Khoo, Artificial neural networks for vibration based inverse parametric identifications: A review. *Applied Soft Computing*, **52**, 203-219, (2017)
12. Z. Rowland, J. Vrbka, Using artificial neural networks for prediction of key indicators of a company in global world. *Proceedings of the 16th International Scientific Conference Globalization and Its Socio-Economic Consequences*, Rajecké Teplice, Slovakia, pp. 1896-1903, (2016)
13. X. Han, L. Wang, Stock Company Comprehensive Assessment Model Based on Kohonen Network. *2008 Second International Conference on Genetic and Evolutionary Computing*, 185-188, (2008)
14. V. Konečný, O. Trenz, Classification of companies with the assistance of self-learning neural networks. *Agricultural Economics*, **56**(2), 51-58, (2010)
15. M. Vochozka, Formation of complex company evaluation method through neural networks based on the example of construction companies' collection. *AD ALTA – Journal of Interdisciplinary Research*, **7**(2), 232-239, (2017)
16. R. Waheed, D. Chang, S. Sarwar, W. Chen, Forest, agriculture, renewable energy, and CO<sub>2</sub> emission. *Journal of Cleaner Production*, **172**, 4231-4238, (2018)
17. J. Bronnmann, F. Asche, The Value of Product Attributes, Brands and Private Labels: An Analysis of Frozen Seafood in Germany. *Journal of Agricultural Economics*, **67**(1), 231-244, (2016)