

Analysis of business companies based on artificial neural networks

Veronika Machová^{1,*} and Marek Vochozka¹

¹Institute of Technology and Business in České Budějovice, School of Expertness and Valuation, Okružní 517/10, 37001 České Budějovice, Czech Republic

Abstract. Business companies have many kinds of products that they sell to other businesses, consumers, etc. They are a driving force of economies, especially in developing countries. The aim of this article is to analyse business companies in the Czech Republic using artificial neural networks and subsequently to estimate the development of this branch of the national economy. An analysis is performed to create a significant number of clusters of businesses. An analysis of the most significant clusters is also carried out. The result can be generalized and we can predict the number of companies that will be creditworthy or bankrupt in the following period. This makes it possible to estimate not only the overall growth or decline of business companies in the Czech Republic, but also to estimate the structure of the companies in terms of their size, turnover or volume of sales.

Key words: artificial neural networks, cluster analysis, business companies, analysis, prediction

1 Introduction

The primary strategic objective of every company should always be to increase its value. The value of a company is very important. There is a need not only to increase the wealth of individual shareholders but also to support the interests of other stakeholders [1]. Therefore, it is very important to analyse and predict the financial sustainability of the company. According to Grant [2], there are many different methods to this end. However, some methods are preferred, especially due to their ease of use, flexibility and versatility. However, there are also methods which are difficult to use and require skills that many analysts lack. Skalický and Puchýř [3] argue that one of the ways to analyse the potential risks arising from the company's financial data and how to oversee the financial health of a company is financial analysis. In advanced market economies, financial analysis has a long tradition. Among company assessment methods, we can also include benchmarking based on non-financial indicators. These indicators should predict the company's future potential and also evaluate its performance [4].

Economic Value Added (EVA) is another indicator that will allow us to uncover the success of a company on the market [5]. The importance of this indicator is increasing as it is used in practical management of a company as a motivation factor for managers and in analysing benchmarking [6]. It is also possible to evaluate the company, for example,

* Corresponding author: machova@mail.vstecb.cz

according to bankruptcy and creditworthiness models [7]. Companies can be evaluated by other methods, such as using value generators, to allow more accurate measurement of the efficiency of the processes and elements that generate the value. According to Valášková et al. [8], these approaches are referred to as value factors. Value generators can be defined as factors that contribute to value creation in a market evaluation of a company. Knowledge is necessary for involved parties to determine future and current internal values of shares [9]. Value factors allow to reflect the environment in which business entities operate and are able to measure the performance of the company and the value of its corporate strategy [10].

Another method for evaluating companies is by using neural networks, which can also be applied in business practice. The advantage of artificial neural networks is their ability to predict future periods. Sánchez and Melin [11] state that neural networks have a huge application in many different areas. Artificial neural networks have innumerable advantages over conventional methods. They allow you to analyse complicated patterns quickly and with high precision and, according to Santin [12], they are flexible in their own use. The disadvantage of these networks is the need for large sample data, since many test observations are needed to produce such data, however this is very complicated for users [13].

Kohonen networks are models of neural networks that can be used to configure data sets into different categories. The data is configured so that the data within one category tends to be the same and the data in different categories are different. A large number of experimental results shows that Kohonen's networks are very effective for business valuation [14]. According to Horák, Vochozka and Machová [15], the Kohonen network consists of an input layer that is connected with the output layer and is taught without a teacher, by self-organization. It can be said that the network is widely used because it is an alternative network that is usable for most neural network calculations. It is primarily used for managing photos, videos, audio, application security, speech processing, and high-dimensional data projection on lower-dimensional data [16].

In this paper, business companies will be analysed by Kohonen Networks. Businesses companies are companies that have many kinds of products that are sold for the needs of other businesses, consumers, etc. According to Dyhdalewicz [17], business companies keep a store, buy a huge amount of products, and buy and supply goods to customers. According to Doganay and Kocsoy [18], business companies are the driving force of economies, especially in developing countries. These business companies are able to connect sellers with buyers and earn revenue through sales commissions.

The aim of the paper is to perform a cluster analysis of businesses operating in trade.

2 Data and methods

For the purposes of this contribution, a set of data on 11,604 retail and wholesale businesses in 2016 will be created. These will be businesses with the main economic activity classified in Section G of economic classification of CZ NACE activities. The data set will be generated from the Albertina database. The complete financial statements of these companies will be available (all financial information is in CZK thousands). For the analysis, only some items will be used:

1. This is information about the size of the business, namely the volume of assets (fixed, current, other).
2. Fixed assets: the item describes the degree of involvement of fixed assets and gives an idea of the level of technology used in production. The higher the share of fixed assets, the higher the production automation can be expected.
3. Current assets: represent the amount of funds, receivables and inventory in the business. It is in most cases the so-called working capital, i.e. those asset

components that change their substance in less than one year and which are directly involved in the core business activity.

4. Equity: reflect the level of business risk for the business owners. This is the amount of funds the owners may lose in the event of bankruptcy.
5. External capital: This is capital that does not have influence on the company's business management. In a way, it reflects the external view of the potential success of the business.
6. Value Added: This is the value added by the company by its core business activity to basic input - material.
7. Operating income: Provides information on how the business is successful in its core business activity.
8. ROE: return on equity means an appreciation of the equity of a business, that is, the capital provided by the business owners.
9. Profit before tax: describes the success or failure of a company in its entire business activity.

These are not generators of value but rather characteristics that may have, albeit indirect, influence on the performance of a business.

The file will then be subjected to cluster analysis, specifically using Kohonen networks. For cluster analysis, Dell's Statistica software in version 12 will be used. Data Mining will be used as a specific Neural Network tool. Here we select neural networks without a teacher – Kohonen's networks. We specify the data for analysis. In all cases, these are continuous predictors. The file will be divided into three parts:

1. Training dataset: represents 70% of companies in the set. The Kohonen network will be created based on this dataset.
2. Testing dataset: this is 15% of the companies of the original set. Using this data set, we verify the parameters of the Kohonen network.
3. Validation dataset: this will also involve 15% of companies in the file. Using this dataset, we will also verify the Kohonen network to see whether it is or is not usable.

We determine the topological length and the topological width of the Kohonen map at 10. The number of iterations of the calculation will be set at 100,000. However, I shall recall that the level of error is decisive. Unless there is any iteration to improve the Kohonen network parameters, training will be terminated before the 100,000 iteration is completed. If the network parameters are improved even at 100,000 iterations, we need to repeat the process and set the higher value of the required iterations to make sure the result is the best possible.

3 Results

In accordance with the methodology of the contribution, the data was divided into three sets – training, testing and validation. The descriptive characteristics of all three sets and the original data set are given in Table 1.

Table 1. Descriptive characteristics of the data file

Samples	Total assets	Fixed assets	Current assets	Equity	External capital	Outputs	Value added	Operating income	ROE	Profit before tax
Minimum (Training)	-1564	-10744.0	-2071	-254305	-533106	-4956.0	-108795	-229908	-644.750	-35843.0
Maximum (Training)	1071267	804579.0	964246	911000	891328	574445.0	301525	234050	841.667	84791.0
Average (Training)	44408	11089.6	32830	20201	23898	3073.5	10335	3387	0.126	49.1
Standard deviation (Training)	78297	32174.8	59367	47471	49033	14845.0	21303	10589	13.297	1293.0
Minimum (Testing)	-8573	-65415.0	-15357	-475270	-8763	-2427.0	-42902	-227728	-871.500	-3998.0
Maximum (Testing)	1067016	430553.0	1039201	505227	901245	258720.0	388474	119588	52.233	44805.0
Average (Testing)	44690	11486.0	32662	18838	25515	2823.1	10146	3040	-0.623	90.2
Standard deviation (Testing)	82682	30419.8	64151	43633	57608	12059.3	21374	10846	22.279	1465.8
Minimum (Validation)	1	-10744.0	-5335	-298827	0	-6242.0	-293974	-84319	-216.000	-25100.0
Maximum (Validation)	1532330	996822.0	771988	1317026	653336	513452.0	407995	197614	18.272	26690.0
Average (Validation)	46190	12539.0	33121	21068	24756	3556.3	10066	3260	0.052	47.7
Standard deviation (Validation)	84300	34684.9	64020	43782	54274	14821.9	21327	9059	21.188	2621.3
Minimum (Overall)	-8573	-65415.0	-15357	-475270	-533106	-6242.0	-293974	-229908	-871.500	-35843.0
Maximum (Overall)	1532330	996822.0	1039201	1317026	901245	574445.0	407995	234050	841.667	84791.0
Average (Overall)	44718	11366.5	32848	20127	24269	3108.3	10267	3316	0.002	55.1
Standard deviation (Overall)	81285	34749.3	60024	48834	50748	15255.5	21768	10592	14.235	1295.5

Source: Authors.

Ideally, if the characteristics of the subsets of data were approximately the same. Since businesses have been randomly distributed, this is not the case. However, the random distribution of data may not be a defect in the result.

Based on the assignment, the Kohonen network was calculated. It is also referred to as SOFM 10-100 in the text. The companies were divided into individual clusters within the 10 x 10 topological grid. The frequency of individual clusters is evident from the three-dimensional graph in figure 1. It is evident from the figure that the largest number of enterprises in individual fields of the topological grid is in the position (1, 10) then in position (2, 10). The following ones are (1, 9), (2, 9) and (2, 8). Overall, we can say that the largest number of the 11,604 companies surveyed is in the sector (1, 7), (4, 7) and (1, 10), (4, 10).

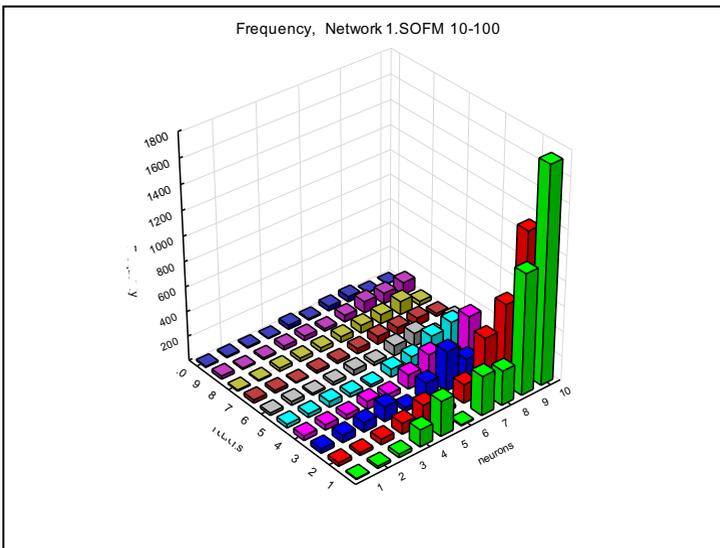


Fig. 1. Graph of business frequencies in topological grid
 Source: Authors.

The results are shown more specifically in Table 2.

Table 2. SOFM 10-100 Network Frequency Table

	1	2	3	4	5	6	7	8	9	10
1	8	19	29	137	281	10	316	278	969	1726
2	29	24	45	88	150	26	151	446	636	1138
3	37	60	78	121	31	148	320	181	97	283
4	38	42	40	64	40	116	201	137	354	60
5	30	30	46	31	20	66	88	178	214	1
6	21	30	16	25	43	28	83	107	15	79
7	28	23	10	22	22	44	65	61	63	25
8	10	10	27	29	34	48	63	75	108	34
9	26	12	16	34	37	22	55	93	87	101
10	9	11	11	6	35	11	37	44	4	3

Source: Authors.

It also follows from the table that it is advisable to focus on four clusters, namely (1, 10), (2, 10), (1, 9) and (2, 9), possibly the fifth cluster (2, 8). This is a very interesting set of data representing more than one third of all companies in the file. Dominance is more evident from the Kohonen graph in Figure 2.

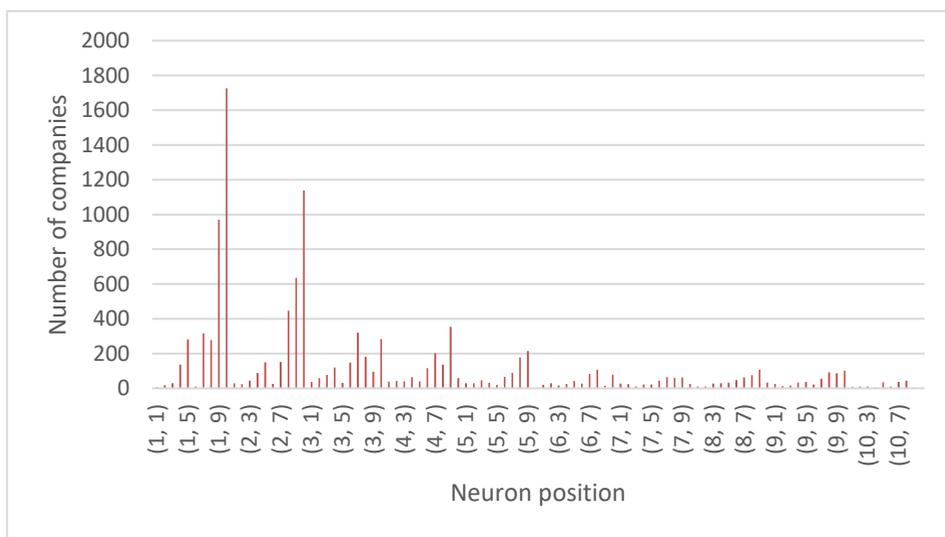


Fig 2. Kohonen graph for selected cells of the SOFM 10-100 network

Source: Authors.

While the data set is rather heterogeneous, cluster analysis suggests that there are characteristics in which businesses are similar.

Cluster (1, 10)

The cluster (1, 10) of the topological grid is occupied by 1,726 companies. It is the largest set of data from partial clusters. From the point of view of input parameters, we can derive the following average values:

1. Total assets: 5,740.1 mil. CZK.
2. Fixed assets: 948.2 mil. CZK.

3. Current assets: 4,705.4 mil. CZK.
4. Equity: 1,174.9 mil. CZK.
5. External capital: 4,520.5 mil. CZK.
6. Outputs: 206.3 mil. CZK.
7. Value added: 1,109.3 tis. CZK.
8. Operating income: 175.2 mil. CZK.
9. ROE: 0.19 %.
10. Profit before tax: 2.3 mil. CZK.

From the average values, we can derive a typical company for this cluster of businesses. It is a small company whose operating income is minimal, 101 thousand CZK per year. It generates a pre-tax profit on average of only 1.3 thousand CZK. Nevertheless, it has assets of CZK 3.3 million. However, the assets are financed for the most part by external capital. In particular, such a company has more than CZK 2.6 million of external capital. Value added is also not high. It is slightly over 680,000 CZK. Nevertheless, the existence of the business is not jeopardized. Based on the nature of its business activity, the company holds a relatively large volume of current assets (probably goods) in excess of CZK 2.7 million.

When it comes to a homogeneous cluster of businesses, we can verify the validity of basic logical links. The relationship between total assets and the amount of operating income appears interesting (see Figure 3 below).

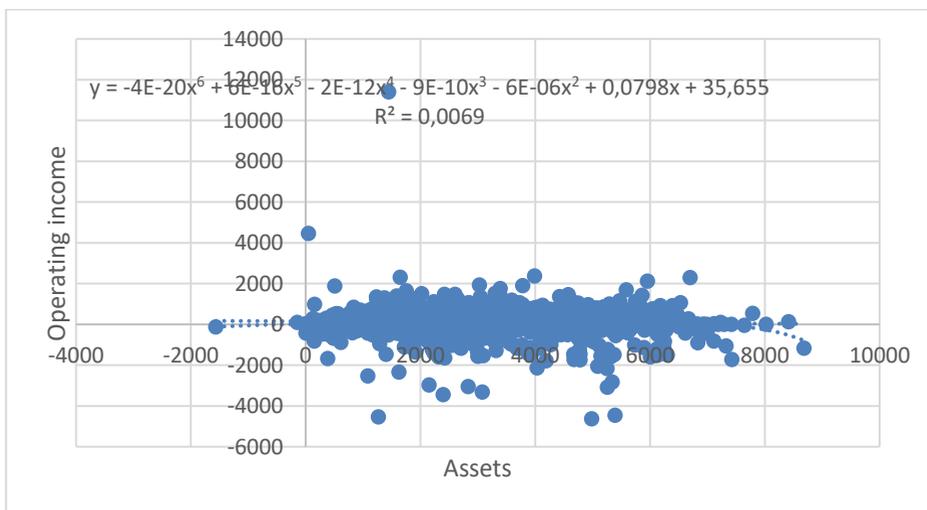


Fig. 3. The relationship between the total assets and the operating income of the cluster (1, 10)
 Source: Authors.

Total assets and operating income are correlated to more than 0.08. Therefore, there can be no dependence of both variables. Yet the data is intersected by a regression curve, a 6th order polynomial. Based on this, it is possible to estimate the development of the operating profit on the basis of the size of the company determined by the volume of total assets.

Figure 4 provides a relationship between fixed assets and operating income.

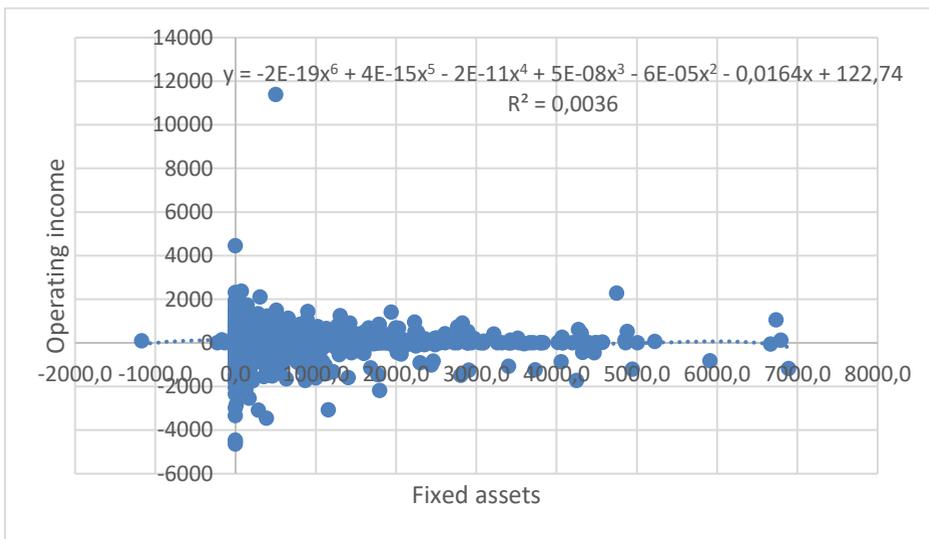


Fig. 4. The relationship of fixed assets and operating profit of cluster (1, 10)
Source: Authors.

The figure shows that there is virtually no dependency between the data. The correlation coefficient is 0.06. By default, the volume of fixed assets represents the degree of automation in the enterprise. In the case of trade business companies, it rather provides information about the ownership of the factors of production – namely, one – the real estate in which the business is located, if it is needed to carry out the business activities. A number of business transactions can go hand in hand with real estate – such as the operation of e-shops. Figure 4 illustrates that trade business enterprises often do not need long-term assets for their business, possibly only in the minimum amount (for example, in the form of a car).

Figure 5 thus presents the relationship between external capital and operating income.

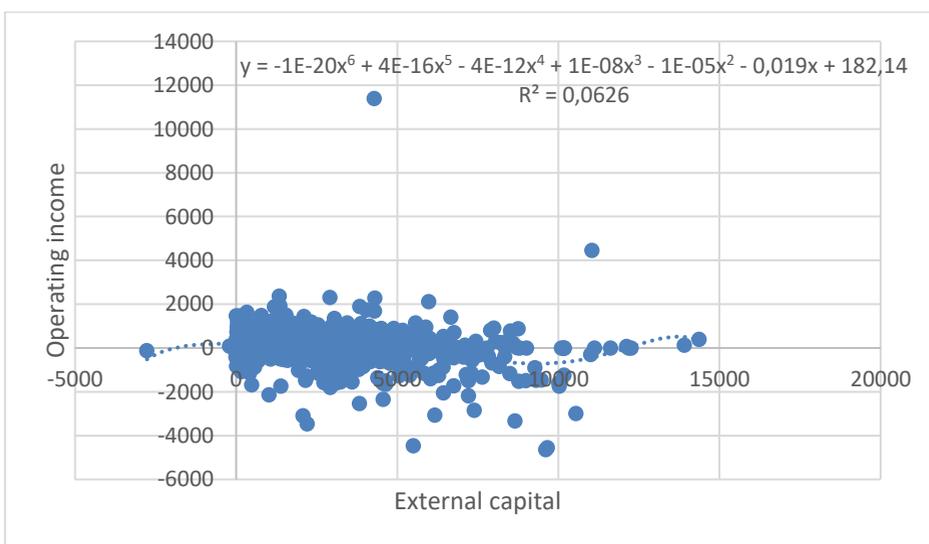


Fig. 5. The relationship between external capital and the operating income of cluster (1, 10)
Source: Authors.

The relationship can show us the effect of a financial leverage, but only assuming that there will be a relatively high correlation between the two variables. However, it's only 0.25. Yet, the graph points are intersected by a 6th order polynomial. However, we can see from the graph in the figure that businesses are in unhealthy debt. Additional external capital does not increase the return on equity. Both variables are indifferent and point to a poor capital structure and perhaps even unnecessary assets in the enterprise. Companies in this cluster should reduce the involvement of external debt.

Cluster (2, 10)

The cluster (2, 10) of the topological grid occupies the imaginary second position in frequency. There are 1,138 companies. From the point of view of the total input parameters we can deduce a typical representative of this cluster:

11. Total assets: 6 668.9 mil. CZK.
12. Fixed assets: 1 364.9 mil. CZK.
13. Current assets: 5 212 mil. CZK.
14. Equity: 2 534.3 mil. CZK.
15. External capital: 4 083 mil. CZK.
16. Output: 344.2 mil. CZK.
17. Value added: 3 667.7 mil. CZK.
18. Operating income: 976 mil. CZK.
19. ROE: 0.81 %.
20. Profit before tax: 20.4 mil. CZK.

A typical company of this cluster is more than 1.75 times larger than the average company of cluster (1, 10), meaning from the point of view of the volume of total assets. It is an enterprise that employs more than 5.8 million CZK worth of assets. The company owns fixed assets in the amount of almost CZK 1.2 million. At the same time, it manages current assets in excess of CZK 4.5 million. Its assets are funded by more than CZK 2.2 million from its own resources, with the remaining amount, i.e. almost CZK 3.6 million, coming from external capital. It achieves an operating income over 850 thousand CZK, which represents almost 18 thousand CZK profit before tax. In addition to the realized profit, a certain shift in the company of this cluster compared to the cluster (1, 10) is a lower indebtedness rate. The companies of cluster (1, 10) had a debt ratio of almost 80%. On the other hand, businesses of cluster (2, 10) have a debt ratio slightly above 62%. Yes, the high level of indebtedness is due to the nature of the business activities. Nevertheless, we can say that companies of cluster (2, 10) are closer to the optimum (even with a larger sum of money than companies of cluster (1, 10)). Even in this case, it is certainly interesting to explore the relationship of some variables.

Figure 6 shows the relationship between total assets and operating income.

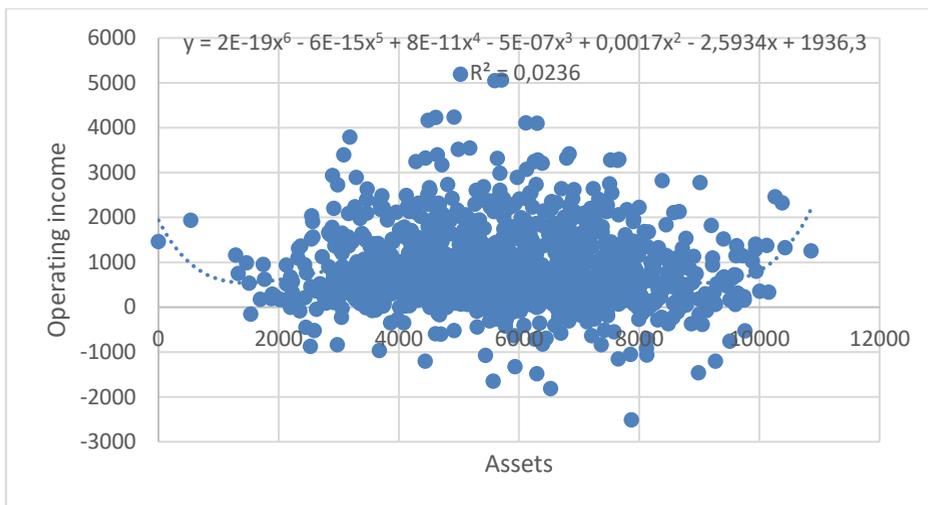


Fig. 6. The relationship between the total assets and the operating profit of cluster (2, 10)
 Source: Authors.

The correlation coefficient of the two variables is 0.15. Also, we can't talk about a dependency between the two variables. However, in this cluster, we can observe the trend in the fact that a higher volume of assets implies a higher operating income. The trend-line has a 6th order polynomial again.

Figure 7 illustrates the relationship between fixed assets and operating income.

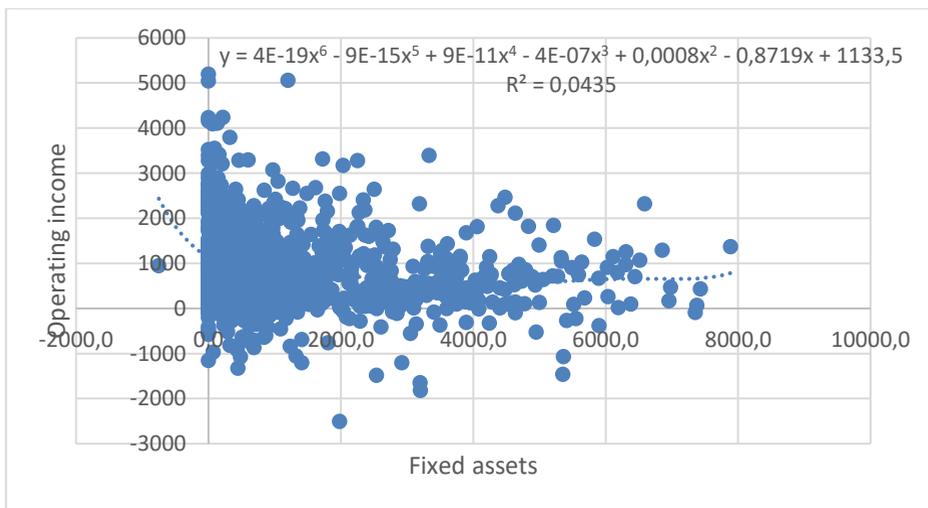


Fig. 7. Relationship between fixed assets and the operating income of cluster (2, 10)
 Source: Authors.

For the relationship between fixed assets and operating income, the correlation coefficient is 0.21. While not a statistically significant dependency, we can already say that the higher volume of fixed assets has a positive effect on the generation of operating income.

Figure 8 provides a relationship of external capital and operating income.

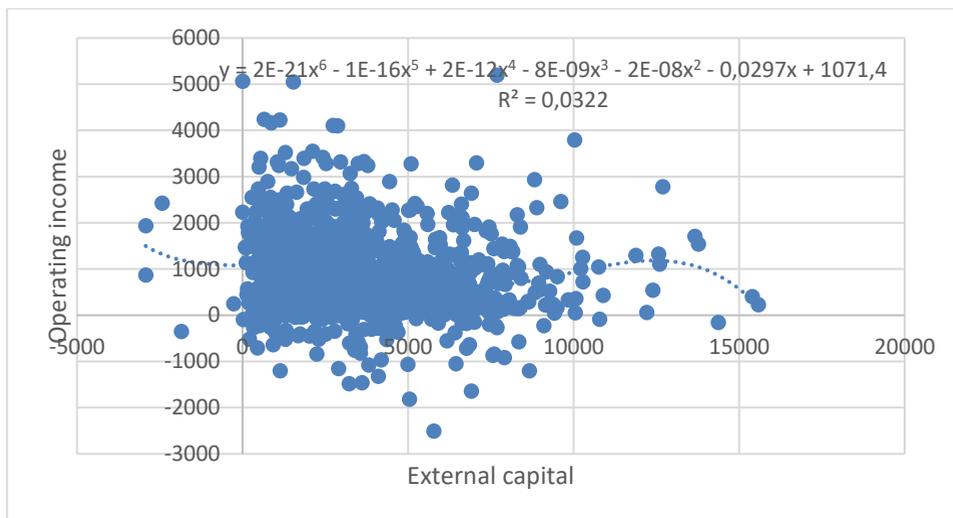


Fig. 8. The relationship of external capital and the operating income of cluster (2, 10)
Source: Authors.

Even in this case, the correlation coefficient is not high, namely 0.18. However, even in this case, we can infer that a higher debt ratio does not mean a higher operating income. As in the case of cluster (1, 10), it is advisable to reduce the volume of external capital involved.

4 Conclusion

The aim of the paper was to perform a cluster analysis of businesses operating in trade. The aim of the paper was fulfilled.

Cluster analysis was performed using neural networks without a teacher – Kohonen networks. Based on cluster analysis, companies were divided into clusters in the Kohonen map (10 x 10 clusters). Some clusters are significant in terms of number of businesses. The most represented are clusters (1, 10) and (2, 10). These clusters have therefore been subjected to further analysis. In the case of both clusters, the typical representatives of these clusters were defined. These varied by their size and volume of operating income generated. The companies of both clusters were on the borderline of profit margins in 2016. The average operating income for one company of cluster (1, 10) was approximately 1.3 thousand CZK, in the case of cluster (2, 10) it was 18 thousand CZK. The companies of the two clusters differ far more in the amount of the operating profit. Companies of cluster (1, 10) make a profit of 101 thousand CZK. Companies of cluster (2, 10) reach 850 thousand CZK in profit. Businesses of both clusters are highly indebted. This can be partly related to the field in which they operate and partly to the non-rational decision about the involvement of external debt. In the case of debt, companies of cluster (2, 10) are better off, with external capital accounting for 61% of all capital. On the other hand, the indebtedness of companies in (1, 10) is almost 80%. According to the volume of fixed assets, it is possible to conclude, in particular, for companies of cluster (1, 10) that they do not operate in their own premises, but in rented premises or on the Internet. From the analysis we can make two very specific conclusions:

1. A larger company (or a company with more assets) generates, on average, a higher operating income.

2. In the case of a larger company, the financial leverage is more positive, i.e. with the growth of the volume of the foreign capital, the operating income and the profitability of the equity increase.

References

1. R. R. Gamayuni, The Effect Of Intangible Asset, Financial Performance and Financial Policies On The Firm Value. *International Journal of Scientific & Technology*, **4**(1), 202-212, (2015)
2. D. Grant, Business analysis techniques in business reengineering. *Business Process Management Journal*, **22**(1), 75-88, (2016)
3. M. Skalický, B. Puchýř, Use of financial analysis for elimination of risk factors in the construction company. *Political Sciences, Law, Finance, Economics and Tourism Conference Proceedings*, Albena, Bulgaria, 669-676, (2016)
4. M. Vochozka, Z. Rowland, J. Vrbka, Financial analysis of an average transport company in the Czech Republic. *Nase More*, **63**(3), 227-236, (2016)
5. I. Ismail, The ability of EVA (Economic Value Added) attributes in predicting company performance. *African Journal of Business Management*, **5**(12), 4993-5000, (2011)
6. J. Vrbka, Modeling the future development of top company indicator EVA Equity using indicator breakdown and sensitivity analysis on example of a specific company. *Proceedings of the 5th International Conference Innovation Management, Entrepreneurship and Sustainability*, Prague, Czech Republic, 1095-1106, (2017)
7. D. Alaminos, A. Del Castillo, M. A. Fernández, G. Ponti, A Global Model for Bankruptcy Prediction. *PLOS ONE*, **11**(11), (2016)
8. K. Valášková, T. Kliešтик, L. Švábová, P. Adamko, Financial risk measurement and prediction modelling for sustainable development of business entities using regression analysis. *Sustainability*, **10**(7), 2144, (2018)
9. E. Vítková, J. Chovancová, J. VESELÝ, Value Driver and Its Impact on Operational Profit in Construction Company. *Procedia Computer Science*, **121**, 364-369, (2017)
10. M. Strnadová, M. Karas, M. Režňáková, Value drivers of Processing Industry Enterprises in the Czech Republic in the Years 2007 -2011. *Trends Economics and Management*, **7**(13), 91-99, (2013)
11. D. Sánchez, P. Melin, Modular Neural Networks for Time Series Prediction Using Type-1 Fuzzy Logic Integration. *Design of Intelligent Systems Based on Fuzzy Logic, Neural Networks and Nature-Inspired Optimization*, **601**, 141-145, (2015)
12. 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)
13. 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)
14. X. Han, L. Wang, Stock Company Comprehensive Assessment Model Based on Kohonen Network. *2008 Second International Conference on Genetic and Evolutionary Computing*, Jingzhou, China, 185-188, (2008)
15. J. Horák, M. Vochozka, V. Machová, Evaluation of transport and forwarding companies by means neural networks. *Proceedings of the International Scientific*

Conference for Ph.D. students and assistants – QUAERE, Hradec Králové, Czech Republic, 377-385, (2018)

16. 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, 985-995, (2017)
17. A. Dyhdalewicz, Innovation of Trading Companies in Relation to the Content of Annual Reports: Research Results. *Economic and Social Development (ESD 2017)*, 629-638, (2017)
18. M. Doganay, M. Kocsoy, A field study concerning the effects of economic crisis on international trading companies and the measures taken. *African Journal of Business Management*, **55**(22), 9477-9489, (2011)