Logistic Regression or Neural Network? Which Provides Better Results for Retail Loans?

KITTI FODOR
ASSISTANT LECTURER

UNIVERSITY OF MISKOLC
e-mail: kitti.fodor@uni-miskolc.hu
https://orcid.org/0000-0003-1815-8159

SUMMARY

While there is extensive literature on the prediction of corporate bankruptcies, there is little literature on the classification of retail borrowers. This is also true in Hungary. Recognising who is at risk of becoming a bad debtor is not easy. There are several ways to analyse the data, which may yield different results. In this paper, my aim is to predict the default of household loans using logistic regression and neural networks. The question is, which method produces the better results? The analyses show that the neural network model produced the best and most favourable results. The accuracy of the best method was found to be 81.5%.

Keywords: statistics; logistic regression; neural network, loan default
Journal of Economic Literature (JEL) codes: B16; C15; C38; C45; C53
DOI: https://doi.org/10.18096/TMP.2023.01.05

INTRODUCTION

For both corporate and retail lending, it is important for financial institutions to lend to customers with a low risk of non-repayment. Although institutions have their own credit assessment process, they still may not properly select the customers for whom it is appropriate to provide credit.

While there is a great deal of literature on the prediction of corporate bankruptcies, there the classification of retail borrowers has received little attention. In the case of the corporate sector, it has been observed that the range of analytical tools used has steadily increased with the development of methodological possibilities and IT, from the initial univariate analyses to the present day models using neural networks.

This is also true in Hungary, where there is not much scientific work on this topic. Despite the fact that the last crisis was not so long ago, in recent years lending has really taken off, with many people taking out loans as if there were no tomorrow. Recognising potential bad debtors is not easy. We can use different methods, and sometimes the results are not the same.

The aim of my research is to investigate the default of household loans using methods based on multivariate statistical procedures. The different methods and models can help to identify the factors that contribute to someone becoming a defaulting debtor, and to determine which multivariate statistical methods produce the best results.

INTERNATIONAL BANKRUPTCY MODELS: LITERATURE REVIEW

Bankruptcy forecasting research does not yet have a 100-year history. The first attempts were made in the 1930s, but the models in use today only appeared in the 1960s. In the period up to the present day, models and methods have changed a lot, which is also due to advances in analytical capabilities and information technology.

In fact, the initial ‘models'’ were not really models at all: researchers were looking for indicators for which there was a significant difference between bankrupt and surviving companies, comparing these indicators and trying to establish various correlations.

The first univariate analysis was conducted by Beaver. He included 158 companies in his analysis, with equal proportions of bankrupt and surviving companies. His method allowed him to categorise companies with 90% classification accuracy. The disadvantage of the method is that it is a univariate model, so the categorisation is based on a single indicator; if different indicators result in different classifications, the method cannot handle this. This is one of the reasons why this method has not been widely used (Beaver, 1966; Virág, 2004).

The first real model was created by Altman, who built his model on five financial indicators that could predict insolvency with 95% confidence. A few years later, an extended seven-variable model was developed based on this model (Altman, 1968; Virág, 2004). Deakin also used discriminant analysis to predict bankruptcy. He used a sample with 34 cases to test his results. The model had a classification accuracy of 97% (Deakin, 1972). Blum's 1974 model also had a classification accuracy of over 90% (Blum, 1974). Altman's extended version of his five-variable model was developed in 1977, and the new model used a larger sample of 111 cases, with 58 surviving companies. (Altman et al., 1977)

Altman's models were not representative, and the sample included roughly equal proportions of surviving and failing firms. Ohlson conducted the first survey to be considered representative. Ohlson was also the first to use logistic regression in bankruptcy prediction models. The sample he studied included 2163 companies, of which 4.85% went bankrupt. If the P-value in the model exceeds 0.038, the company is considered to be at risk of bankruptcy. The model has an accuracy of approximately 83% (Ohlson, 1980).

The next novelty was the emergence of recursive partitioning algorithms and dates back to the mid-1980s. Among the first adopters of this method were Altman, Frydman and Kao. The classification accuracy of the model was 94%, but there was a significant difference in the correct categorisation between surviving and failed firms (Frydman et al., 1985)

The next big step was the emergence of neural networks, which dates back to the 1990s. The first application of neural networks can be attributed to Odom and Sharda. Their model was based on the variables used by Altman in 1968. The sample consisted of 129 companies. For the training sample, the classification was perfect, thus outperforming the results obtained by discriminant analysis. For the test sample, the classification accuracy of 82% significantly exceeded the results obtained by discriminant analysis (Odom & Sharda, 1990)

Tam and Kiang's analyses were carried out for banks, with the neural network performing best over a one-year time horizon, but logistic regression performing best over a two-year time horizon (Virág & Kristóf, 2005). Coats and Fant compared the performance of discriminant analysis with a neural network and came to similar conclusions. In the second half of the 1990s, Olmeda and Fernandez processed data from Spanish banks. Their research was carried out using all the models mentioned above, of which the neural network proved to be the best, with a classification accuracy of 82.4% (Olmeda & Fernandez, 1997). Zhang et al. (1999) compared the neural network with logistic regression. The former achieved a classification accuracy of 88.2% and the latter 78.6%.

Overall, it can be concluded that, among the different methods of analysis, neural networks have basically produced the best results.

METHODS OF PREDICTING BANKRUPTCY

For bankruptcy forecasting, the following methods are widely used:

- discriminant analysis
- logistic regression
- decision tree
- neural network

For this study I used logistic regression and neural network.

Logistic regression

In logistic regression, the goal is to classify observation units into predefined groups of dependent variables. In this case, the dependent variable has two categories, so I applied binomial logistic regression. In logistic regression, the analysis is based on the ‘odds’, which determine the probability of the default. The odds can be expressed by the following formula:

$$odds_x = \frac{P_x}{1 - P_x}.$$

In the logistic regression, we assume that the logarithm of the odds can be defined as a linear function of the independent variables, which can be written as follows:

$$\ln(odds_x) = logit(P_x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p \,.$$

The other central element of the analysis is the so-called cut point value. This value can be chosen arbitrarily, but it is important to keep in mind that the losses resulting from a false classification are kept to a minimum (Hajdu, 2003, 2018; Sajtos & Mitev, 2007; Malhotra, 2008; Varga & Szilágyi, 2011).

Neural network

The best performing computer today is still the human brain. Neurons, information-processing units, help to perform tasks. “Neural networks, or more precisely artificial neural networks, are information processing paradigms inspired by the highly interconnected parallel processing structures and processes of the mammalian brain. In essence, neural networks are mathematical models that operate on the basis of certain information processing principles of biological nervous systems and are therefore capable of adaptive learning.” (Ketskeméty et al., 2011, p. 394).

In my analyses, I used the Multi-Layer Perceptron (MLP) method, which extends the simple perceptron with hidden layers that are placed between the input and output layers, improving learning performance. Information can flow between the layers with and without feedback. The best known is the back propagation network, where the error propagates backwards, continuously shaping the weights (Ketskeméty et al., 2011).

Neural networks have several advantages:

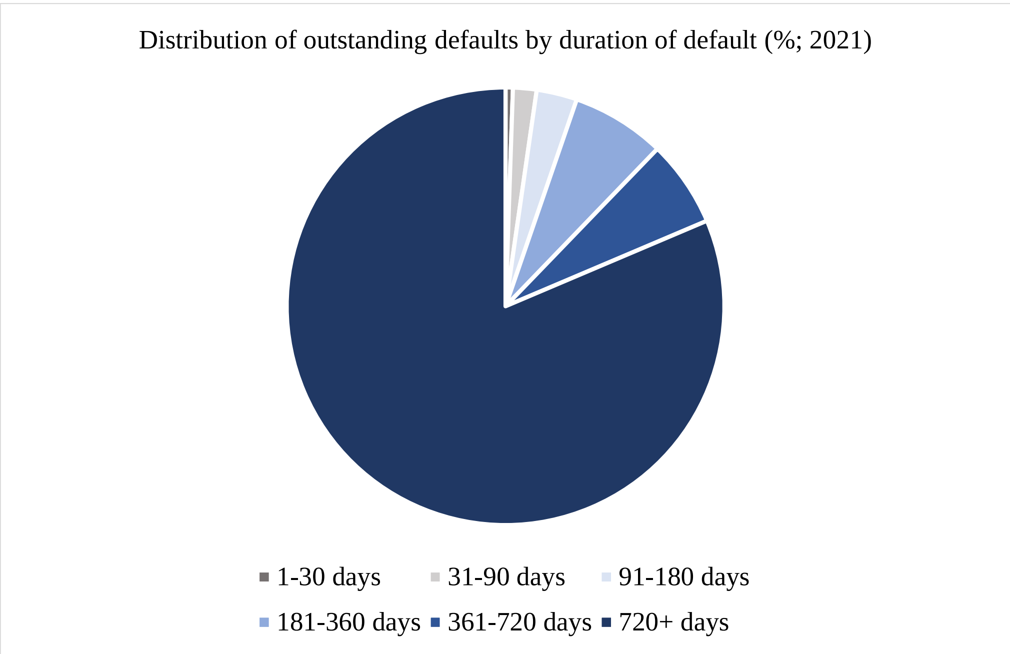
- they can handle nonlinearity;
- they have no problem with missing data;
- they can handle large numbers of variables and elements (Kristóf, 2002).

To evaluate and compare the models, I used the classification matrix, ROC curve and Gini coefficient. For the AUC value calculated from the ROC curve, if the value is around 80–90%, it is considered to be outstanding. Also for the Gini coefficient, a value between 80–90% indicates a very good predictive model.

DATABASE AND RESULTS

In Hungary, information on household creditors is kept by the Central Credit Information System, or KHR, which helps banks to share information on creditors, assist in credit assessment and reduce the risk of over-indebtedness. The KHR maintains a so-called complete list, i.e., customers who meet their obligations on time are also included in the register.

The trend in defaults for 2021 shows that the number of defaults steadily decreased over the year, with the number of outstanding debts falling by 13.9% from January to December. The trend in outstanding debts has also been influenced by legislative changes, such as the gradual reduction of the moratorium on repayments.



Source: Own editing based on KHR data

Figure 1. Distribution of outstanding defaults by duration of default

Figure 1 shows the distribution of outstanding default by duration. In terms of duration of defaults, 12.21% of the outstanding defaults have been outstanding for up to one year, 6.4% for less than 720 days and a significant proportion, 81.39%, for more than almost 2 years (KHR Annual Information, 2021).

The necessary database for the analyses was provided by BISZ Zrt. The data were extracted on 30 September 2021, so the database contains the persons registered on that date. A unit in the database represents one loan transaction, so there may be persons in the database who are listed more than once with different loan transactions. Overall, on that date, the register contained 10,767,452 credit transactions and 21 variables. In addition to the original variables, I added more variables to the database. The original variables are:

- Anonymous identifier of the consumer
- Type of the consumer
- Age of the consumer (in 2021)
- Gender of the consumer
- Distorted agreement identifier
- Type of agreement
- Status of agreement
- Date of the agreement
- Expiry date of the agreement
- Amount and currency of the agreement
- Amount and currency of principal debt outstanding
- Information on regular repayments (amount and currency)
- Amount and currency of the default
- Status of the default
- Date the default occurred
- Date the default was terminated
- Residence of consumer (on the district level of the country)

Added variables are:

- Default (yes or no)
- Loan maturity (difference between the date and expiry date of the agreement; in months)
- Repayment amount as a percentage of agreement amount (repayment amount/amount of agreement)
- Age of the consumer at the time of borrowing the loan

For the analysis the relevant variables are the default, gender, loan maturity, age of the consumer at the time of borrowing the loan, and repayment amount as a percentage of agreement amount.

Before starting the analyses, the first step was to clean the database and narrow it down to the research objectives; after that I had 2,887,470 cases in the database. For the analysis I used 2 database with 500 cases. For the sampling I used a random numbers generator. For the first sample, I used simple random sampling. This is a type of representative sampling. For the second sample, I also used random sampling, but in this case the proportion of performing and non-performing loans is the same. The second sample type is a good and applied practice in this area.

Empirical research

Recent methods used for bankruptcy prediction include logistic regression and neural networks. I assume that these methods can be used to predict with high accuracy which customers or loan transactions will default.

I classified as default the loan transaction that had a default amount.

To support this statement, I constructed classification models using logistic regression and neural network. To perform the analysis, I used the database provided by KHR and to validate the results, I divided the sample into a training and a test part. The training sample included 70% of the cases.

Logistic regression

First, I performed a logistic regression analysis on the first sample. For the variable selection I used the Backward method. Of the available explanatory variables, only the ratio of the repayment to the agreement amount was found to be significant. The Omnibus test ($p<0.001$) and the Hosmer and Lemeshow goodness-of-fit test ($p=0.212$) showed a reliable model with a good fit. The categorisation (Table 2) of good-performing loans is much more likely to be correct. This may be due to the predominance of good-performing loans in the sample, i.e., the sample composition is unfavourable for analysis. The solution can be using the second sample.

I repeated the analysis again, this time on the second sample. The Omnibus test ($p<0.001$) and the Hosmer and Lemeshow goodness-of-fit test ($p=0.105$) showed a reliable model with a good fit. The variable that was found to be significant in the previous case was also found to be significant in this case ($p<0.001$), and the model was extended to include loan maturity ($p<0.001$) as shown in Table 1.

Table 1

Significant variables in the logistic regression models

Type of the sample		Variable	Wald	Sig.	Exp(B)
I	Test	the ratio of the repayment to the agreement amount	13.249	<0.001	1.025
	Training	the ratio of the repayment to the agreement amount	49.181	<0.001	1.030
II	Test	the ratio of the repayment to the agreement amount	17.529	<0.001	1.031
		loan maturity	8.190	<0.001	0.977
	Training	the ratio of the repayment to the agreement amount	29.528	<0.001	1.051
		loan maturity	9.418	<0.001	0.982

Source: SPSS output, own editing

The question is, which model is better? I think it is a complex issue. If we look at classification accuracy only (Table 2), the first one is better. However, it should be taken into account that the new sample composition has had a positive effect on the categorisation of non-performing loans, which can be considered as a more favourable result for the analysis. Based on this, the second model is the better.

Table 2

Classification tables

Sample type	Observed		Predicted					
			Default		Percentage Correct	Default		Percentage Correct
			0	1		0	1	
Test	Default	0	141	1	99.3	60	27	69.0
		1	4	4	50.0	16	60	78.9
	Overall Percentage				96.7			73.6
Training	Default	0	315	8	97.5	143	20	87.7
		1	14	13	48.1	36	138	79.3
	Overall Percentage				93.7			83.4
			a. The cut value is 0.3			a. The cut value is 0.4		

Source: SPSS output, own editing

The model equation can be written in the following form:

$$Y = \ln(odds_x) = 0.021 + 1.051x_1 + 0.982x_2$$

after transformation

$$P_{(default)} = \frac{e^{0.021+1.051x_1+0.982x_2}}{1 + e^{0.021+1.051x_1+0.982x_2}},$$

where

x_1: the ratio of the repayment to the agreement amount

x_2: loan maturity.

Neural network

In the case of neural networks, I chose the Multi-Layer Perceptron (MLP) option, which is widely used in bankruptcy prediction. The neural network does not have any preconditions, but there is a risk of over-learning. All of the available variables can be included in the analysis. The algorithm chose the ratio of the repayment to the agreement amount as the most important explanatory variable. The resulting model has a high classification accuracy, with 6.7% of loan transactions misclassified in the case of the training part and 3.2% in the case of the testing part.

I encountered the same problem as in logistic regression. In the classification matrix (Table 3), we can see that the sensitivity value is lower than the specificity value, so the analysis achieves higher accuracy for the classification of good-performing loans. A solution could be the same as in the logistic regression, using a sample with (approximately) equal proportions of performing and non-performing loans.

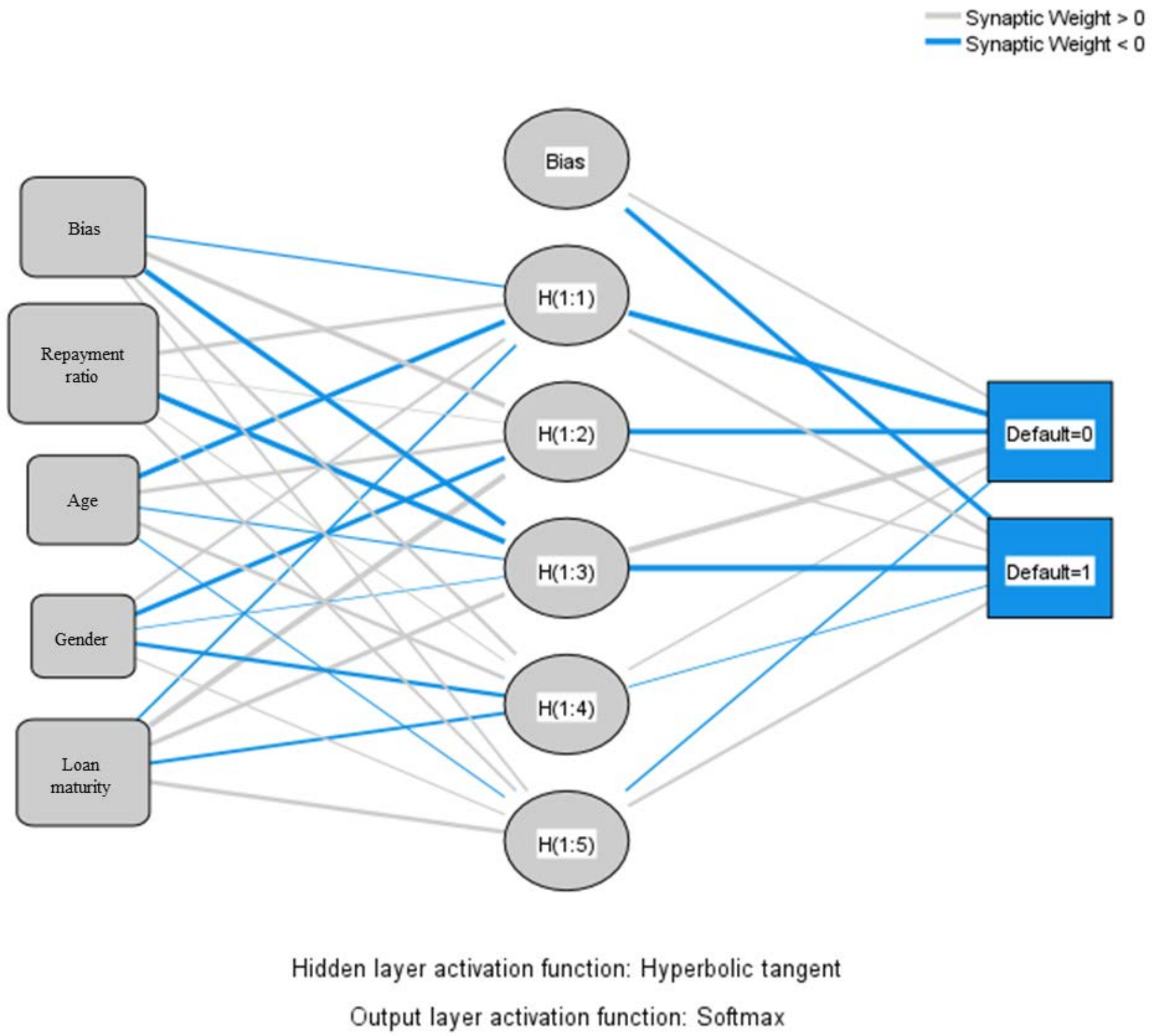
Table 3

Classification table of the neural network model

Sample type	Observed	Predicted		
		0	1	Percent Correct
Training	0	307	9	97.2%
	1	14	12	46.2%
	Overall Percent	93.9%	6.1%	93.3%
Test	0	149	0	100.0%
	1	5	4	44.4%
	Overall Percent	97.5%	2.5%	96.8%

Source: SPSS output, own editing

The second sample produced the neural network shown in Figure 2:



Source: SPSS output, own editing

Figure 2. Neural network model

The Algorithm chose the ratio of repayment to the agreement amount as the most important explanatory variable. Based on the normalised importance of the variables, the hierarchy of the variables is:

- Repayment ratio: 100.0%
- Loan maturity: 66.9%
- Age: 23.0%
- Gender: 7.7%

The resulting model has a high classification accuracy. Although the overall accuracy of the classification has decreased, as can be seen in Table 4, the specificity and sensitivity are approaching each other and the classification of non-performing loans has improved significantly.

Table 4

Classification table of the neural network mode

Sample Type	Observed	Predicted		
		0	1	Percent Correct
Training	0	163	14	92.1%
	1	53	132	71.4%
	Overall Percent	59.7%	40.3%	81.5%
Test	0	73	0	100.0%
	1	19	46	70.8%
	Overall Percent	66.7%	33.3%	86.2%

Source: SPSS output, own editing

Based on this, I think that the second model is the better, and the new sample has a positive effect on the categorisation of non-performing loans.

COMPARISON OF RESULTS, CONCLUSION

I based the models on four explanatory variables, and Table 5 summarises which explanatory variables were found to be significant by the different methods.

Table 5:

Summary of variables used by classification models

Name of the variable	Logistic regression		Neural network	
	I	II	I	II
ratio of repayment	X	X	X	X
loan maturity		X	X	X
age			X	X
gender			X	X

Source: Own editing

Based on the above, it can be concluded that the most significant of the data recorded by the KHR in terms of loan defaults is the ratio of the repayment to the agreement amount.

It can be concluded that the chosen methods can be successfully applied to predict defaults. The accuracy of the classification of the models was high, but there was a significant difference in the classification of each group when the first sample was used. I therefore carried out the analyses on the second sample, where the proportion of performing and non-performing loans is the same. This had a positive effect on the research aim. I also evaluated and compared the models using the AUC value and the Gini coefficient.

Table 6

Evaluation of the models

Method & sample	Accuracy (%)			AUC (%)	Gini (%)
	0	1	∑		
Logistic regression I	97.5	48.1	93.7	87.7	75.4
Logistic regression II	87.7	79.3	83.4	91.2	82.4
Neural network I	97.2	46.2	93.3	90.8	81.6
Neural network II	92.1	71.4	81.5	92.5	85.0

Source: SPSS output, own editing

For the AUC value, a value between 80–90% is considered to be outstanding. In the Table 6 we can see that all models have AUC values significantly above 80%. A similar conclusion can be drawn for the Gini coefficient, where a value above 70% indicates a very strong model. On the basis of classification accuracy, the first logistic regression model is considered to be the best, but on the other two evaluation criteria, the second neural network model is better.

From an application and interpretation point of view, logistic regression is a simpler option, and its performance is barely inferior to that of neural networks, so the best two models are the second logistic regression and the second neural network model.

REFERENCES

Database: provided by Bisz Zrt.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589-609. https://doi.org/10.2307/2978933

Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETATM analysis A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, *1*(1), 29–54. https://doi.org/10.1016/0378-4266(77)90017-6

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, *4*, 71-111. https://doi.org/10.2307/2490171

Blum, M. (1974). Failing company discriminant analysis. *Journal of Accounting Research*, *12*(1), 1-25. https://doi.org/10.2307/2490525

Coats, P. K., & Fant, L. F. (1993). Recognizing financial distress patterns using a neural network tool. *Financial Management*, *22(3)*, 142-155. https://doi.org/10.2307/3665934
Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of Accounting Research*, *10*(1), 167-179. https://doi.org/10.2307/2490225
Frydman, H., Altman, E. I., & Kao, D.-L. (1985). Introducing recursive partitioning for financial classification: The case of financial distress. *The Journal of Finance*, *40*(1), 269–291. https://doi.org/10.1111/j.1540-6261.1985.tb04949.x
Hajdu, O. (2003): *Többváltozós statisztikai számítások [Multivariate statistical calculations]*. Budapest: Központi Statisztikai Hivatal.
Hajdu O. (2018). Többváltozós statisztikai R Open alkalmazások [Multivariate Statistical R Open Applications]. *Statisztikai Szemle, 96*(10), 1021-1047. https://doi.org/10.20311/stat2018.10.hu1021
KHR Annul Information (2021). https://www.bisz.hu/dokumentumtar (May 2023)
Ketskeméty, L., Izsó, L. & Könyves Tóth, E. (2011). *Bevezetés az IBM SPSS Statistics programrendszerbe [Introduction to IBM SPSS Statistics]*. Budapest: Artéria Stúdió Kft.
Kristóf, T. (2002). *A mesterséges neurális hálók a jövőkutatás szolgálatában [Artificial neural networks in Futures Studies]*. Futures Studies Department, Corvinus University of Budapest. https://doi.org/10.13140/RG.2.1.2835.6562
Malhotra, N. K. (2008): *Marketingkutatás [Marketing research]*. Budapest: Akadémiai Kiadó
Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *1990 IJCNN International Joint Conference on Neural Networks*, 163-168. https://doi.org/10.1109/IJCNN.1990.137710
Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, *18*(1), 109-131. https://doi.org/10.2307/2490395
Olmeda, I., & Fernández, E. (1997). Hybrid Classifiers for Financial Multicriteria Decision Making: The Case of Bankruptcy Prediction. *Computational Economics*, *10*(4), 317–335. https://doi.org/10.1023/a:1008668718837
Sajtos, L. & Mitev, A. (2007): *SPSS kutatási és adatelemzési kézikönyv [SPSS research and data analysis handbook]*. Alinea Kiadó.
Varga, B. & Szilágyi, R. (2011). *Quantitative Information Forming Methods.* Nemzeti Tankönyvkiadó. https://www.tankonyvtar.hu/hu/tartalom/tamop425/0049_08_quantitative_information_forming_methods/6127/index.html
Virág, M. (2004): A csődmodellek jellegzetességei és története [Characteristics and history of bankruptcy models]. *Vezetéstudomány, 35*(10), 24-32.
Virág, M., & Kristóf, T. (2005): Az első hazai csődmodell újraszámítása neurális hálók segítségével [Recalculation of the first domestic bankruptcy model using neural networks]. *Közgazdasági Szemle, 52*(2), 144–162.
Zhang, G., Hu, M. & Patuwo, B. (1999): Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, *116*(1), 16–32. https://doi.org/10.1016/S0377-2217(98)00051-4