Mapping the Unified Field: Artificial Intelligence Foundations for the Financial Instruments of Business Planning: Hybrid Data Mining and the Consciousness Model

Dr. Laskai András

PhD-Candidate, University of Sopron, István Széchenyi Management and Organisation Sciences Doctoral School, Sopron, Hungary

Abstract: ***With the continuous development of data mining instruments we acquire increasingly accurate and deep knowledge regarding the processes and network correlations related to business planning that can be considered one of the cornerstones of the world of business. During data mining activity patterns can be derived that provide assistance in the analysis of basic correlations. By the application of data mining a consciousness model can be formulated, which evaluates and analyzes company movements, operations and decision making mechanisms. Data mining is one of the most important current paradigms of advanced business analyses and decision making instruments. (AMANIA-FADLALLA 2017). It is a multidisciplinary approach that applies various techniques in the areas of statistics, machine learning and databases. One of the special branches of data mining inspects the correlations of financial balance sheets and reports as well as their derived indicators. In the case of this method, data mining provides the identification of data patterns in an innovative and ultimately interpretable manner. (PUJARI 2001) Data mining makes it possible for organizations to identify the statistical correlations between performance indicators more easily (ITTNER-LARCKER 2001).***

Keywords: Consciousness model, Data mining, Artificial network, International Business Planning, Financial instruments, Consciousness

1. Introduction

1.1 Introduction of the general professional literature background in the analysis of processes, methods and financial instruments

The information technology inspection of balance sheet analyses, reports and financial analyses commenced in the 1960s, the research of computer auditing standards started in the 1990s, primarily focusing on standards and specifications. (ZHANG-WANG-ZHANG 2011)

In the area of accounting intelligent applications have been used for over three decades (BALDWIN 2006) and for the purpose of the utilization of data mining one of the first business disciplines was the discovery of complexity and system risks. Numerous research projects were conducted in connection with the accounting applications of data mining; however the majority of them focused on a specific field of accounting or data mining technique. (COAKLEY-BROWN 2000, YANG 2006, CALDERON-CHEH 2002, WANG 2010, NGAI 2011.)

The most significant data mining techniques are artificial neural networks, event based argumentation, genetic algorithms, decision trees, association rules, regression, self-organization maps, the K nearest neighbor search as well as the Bayes and fuzzy analyses. (AMANIA-FADLALLA 2017)

All of these data mining techniques serve a special purpose, problem or business need. Various dedicated summary studies have been published regarding the application of data mining and financial as well as accounting applications. The large summary works classify, compare and summarize various data mining methods, algorithms and performance measurements, thus the works by YANG 2006, WANG 2010, NGAI-HU-WONG –CHEN-SUN 2011. At the same time, other larger summary works analyze financial data mining through neural programming methods: COAKLEY-BROWN 2000, CALDERON-CHEH 2002.

1.2 Special data mining processes and methods in the analysis of financial instruments

The authors, Dattilo G-Greco S-Masciari E.-Pontieri L., for the analysis of balance sheet data created the system named DMTool by the combined application of various data mining and classification techniques (DATTILO-GRECO-MASCIARI-PONTIERI 2000). The architecture and main functional viewpoints of their system were expanded to the entirety of the classification and data description process of Italian company balance sheets. This model provides an integrated environment for data management and classification as well as for the analysis of the results. It fundamentally applies clustering and decision tree induction algorithms. (DATTILO-GRECO-MASCIARI-PONTIERI 2000)

According to another method, a parallel simulation program was constructed for a single company, based on the logic of a balance sheet. The basic purpose of the work by Li Zhang-Lu Wang-Jianping Zhang was the discovery of the cross section of balance sheet audit problems arising in the system audits of large corporate groups. The filtering out of unusual data and data analysis were performed according to the rules of correlation analysis, with the application of combined data mining methods and statistical models. (ZHANG-WANG-ZHANG 2011)

According to professional literature analyses, the characteristic of data mining instruments is 82% predictive, 11% descriptive and 7% prescriptive. The analysis of literature shows that neural networks are the most broadly applied technique. They were used by almost half (47%) of all applications. This level of dominance by neural networks may originate from the nature of neural networks, as general problem solving techniques, which are applied in all data mining types and tasks as well as business problems. Regression represented 20% of all applications. Ranked after it with 14% were decision trees, while support vector machines and genetic algorithms were used by 11% of applications. Other less broadly applied techniques include: self-organization maps, the K nearest neighbor search, discriminant analysis association rules, event based argumentation and clustering.

Financial and accounting applications primarily inspected financial performance and analysis. One of the earliest applications of data mining in this area is the work of Callen et al, in which a neural network model was constructed for the forecasting of quarterly accounting revenues. This work compared neural networks with linear time series forecasting models, and described that linear time series forecasting models produced better quarterly revenue forecasts, then an artificial neural network model. The *replicability* of the experiment conducted by Callen et al was difficult because of the absence of an accurate definition of the neural network model. (CALLEN 1996)

1.3 Formulation and general description of the consciousness model

The model is divided into 3 basic interfaces. The first interface is the data request interface. This is where the data is entered for the evaluation. The basic feature of the data request interface is that the data is entered in 10 year periods. Any different data entry results in a non-conscious company condition by the evaluation.

After data entry the data are transferred into the evaluation module, which appears on a separate hidden interface, this is the second interface, the model's "artificial intelligence" (hereinafter referred to as: MMI). The MMI evaluates the correlations in the data. The evaluation module channels the data into 3 evaluation ranges: (i) non-conscious company behavior, (ii) slightly conscious company behavior, and (iii) conscious company behavior. The third interface appears separately as the result. The evaluation module first arranges the entered data into an order of consciousness, then evaluates it in a text, and secondarily supports the evaluation numerically.

The system of conditions serving as the basis for the MMI model are comprised of (i) the abstractions and parameters of the conclusions of theoretical professional literature (ii) the results of the Orbis system test and (iii) the correlations discovered by me. I have already detailed the basics of the first two elements in Figures 3-4. I furthermore data visualized the innovational quality definition used by the Orbis system.

I used this form of the data for testing, for both type of datasets.

Equation 1: Aggregate of innovation data

$$|B|=\begin{bmatrix} \begin{Vmatrix} V_1 & j_1 \\ I_1 & k_1 \end{Vmatrix} & \begin{Vmatrix} V_1 & j_2 \\ I_1 & k_2 \end{Vmatrix} & \begin{Vmatrix} V_1 & j_3 \\ I_1 & k_3 \end{Vmatrix} & & \begin{Vmatrix} V_1 & j_m \\ I_1 & k_m \end{Vmatrix} \\ \begin{Vmatrix} V_2 & j_1 \\ I_2 & k_1 \end{Vmatrix} & \begin{Vmatrix} V_2 & j_2 \\ I_2 & k_2 \end{Vmatrix} & \begin{Vmatrix} V_2 & j_3 \\ I_2 & k_3 \end{Vmatrix} & \cdots & \begin{Vmatrix} V_2 & j_m \\ I_2 & k_m \end{Vmatrix} \\ \begin{Vmatrix} V_3 & j_1 \\ I_3 & k_1 \end{Vmatrix} & \begin{Vmatrix} V_3 & j_2 \\ I_3 & k_2 \end{Vmatrix} & \begin{Vmatrix} V_3 & j_3 \\ I_3 & k_3 \end{Vmatrix} & & \begin{Vmatrix} V_3 & j_m \\ I_3 & k_m \end{Vmatrix} \\ & \vdots & & \ddots & \vdots \\ \begin{Vmatrix} V_m & j_1 \\ I_m & k_1 \end{Vmatrix} & \begin{Vmatrix} V_m & j_2 \\ I_m & k_2 \end{Vmatrix} & \begin{Vmatrix} V_m & j_3 \\ I_m & k_3 \end{Vmatrix} & \cdots & \begin{Vmatrix} V_m & j_m \\ I_m & k_m \end{Vmatrix} \end{bmatrix}$$

where:

$|B|$ Innovation indicators dataset
$V_{(i)}$ International company (1-1000)
$I_{(i)}$ Quality variable (1-21)
$j^{(i)}$ Annual value of the quality variable (1-10)
$k^{(i)}$ Quarterly value of the quality variable (1-10)
Source: Own source. Own editing.
As the result of the test I received the following results:

From the aspect of data related to information technology companies:

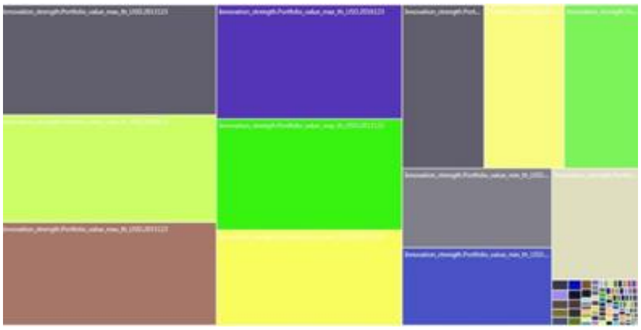

Figure 1: K-centroid cluster data visualization of Infoalldt's innovation indicators.
Source: Own source. Own editing

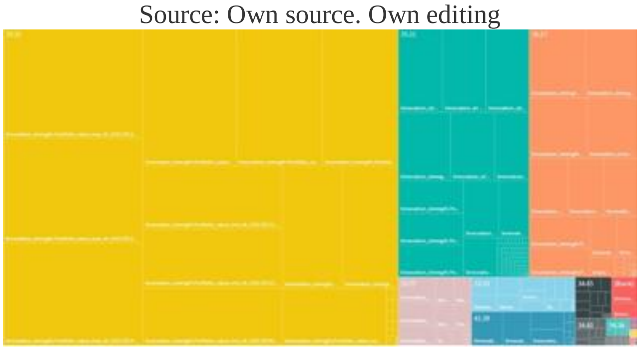

Figure 2: K-centroid cluster data visualization of Infoalldt's innovation indicators.
Source: Own source. Own editing

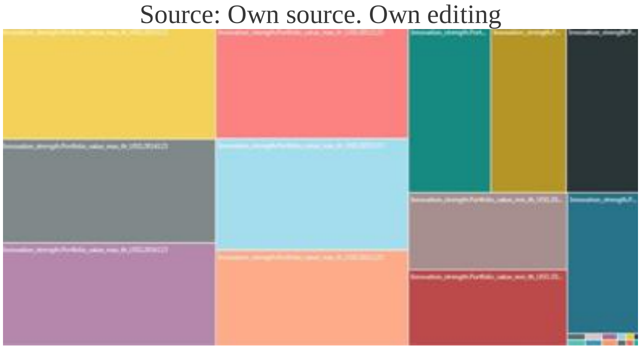

Figure 3: K-centroid cluster data visualization of

NAICS2012 innovation indicators.
Source: Own source. Own editing

Figure 4: K-centroid cluster data visualization of NAICS2012 innovation indicators.
Source: Own source. Own editing

From the aspect of both, the data visualization of K-centroid clustering primarily displayed the Innovation strength-Portfolio value max. For this reason I integrated this value into the consciousness model, as the element showing innovation and defining quality.

I divided the datasets originating from the Orbis system according to the above equation, into derived and non-derived as well as annual and quarterly datasets. A factored all data and subjected them to principal component analysis, then I clustered them.

I primarily took into consideration the elements defining the specific factor, and I primarily filtered the data that were contained 100 percent in every element set.

The factors of the derived values and their components are 40 in the case of both company types. The Figure shows. I indicated the elements in the same factor group with yellow color, which were contained 100 percent in every factor.

Defining elements in the case of derived indicators: a Profit margin, EBITDA margin, Gross margin, P/L for period net income. In the derived portion, the independent factor elements or their parts in at least 75 %: ROE using P/L before tax, Net assets turnover Quarter, Solvency ratio (Asset based), ROCE using net income, ROA using net income.

I determined the factors extracted from the Orbis system relative to the 1,000 element number, which I considered internal values within the model.

The consciousness boundary values described in theoretical professional literature are mostly identical with the results of my analyses. The conclusions in theoretical professional literature have been proven by empirical studies. The important consciousness boundary values, which occur in both datasets: number of employees, gross profit margin, revenue values, liquidity, solvency.

2. Conclusions

2.1 Evaluation of the consciousness model

The consciousness model indicates the specific partial results with colors and pattern. Thus, the color blue is non-conscious, yellow is slightly conscious, and green is conscious in its features. The summary result is indicated in the last line of the model. Summarizing all of these, in the comparison of annual data, from the aspect of Infoalldt, the number of non-conscious companies is 238, that of slightly conscious companies is 1, that of non-conscious companies is 701. From the aspect of NAICS2012, also in the comparison of annual data, 495 companies carry the pattern a conscious company based on the model, the 120 companies the pattern of slightly conscious, and 385 companies the pattern of non-conscious .

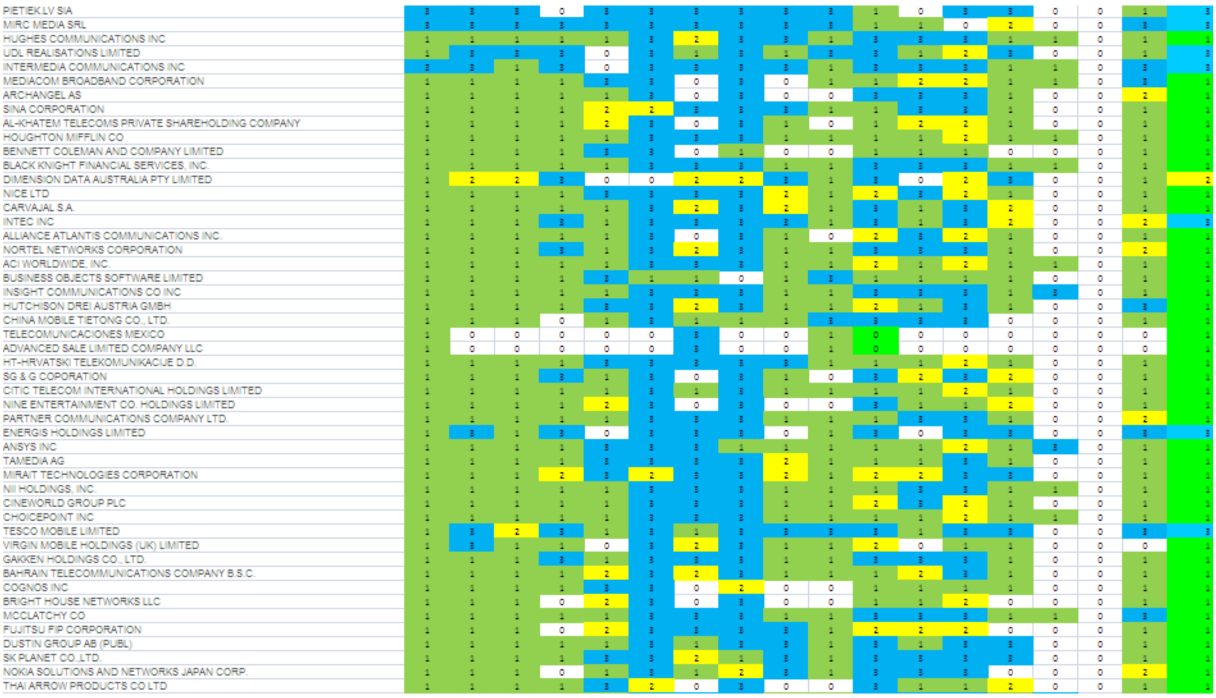

Figure 5: Result patterns of annual data of innovative companies generated by the consciousness model (excerpt).

Source: Own source. Own editing

The two distinguishable patterns in groups, consciousness is an important factor in in the case of the model results of information technology companies as well as mixed companies, which is shown by the similar pattern.

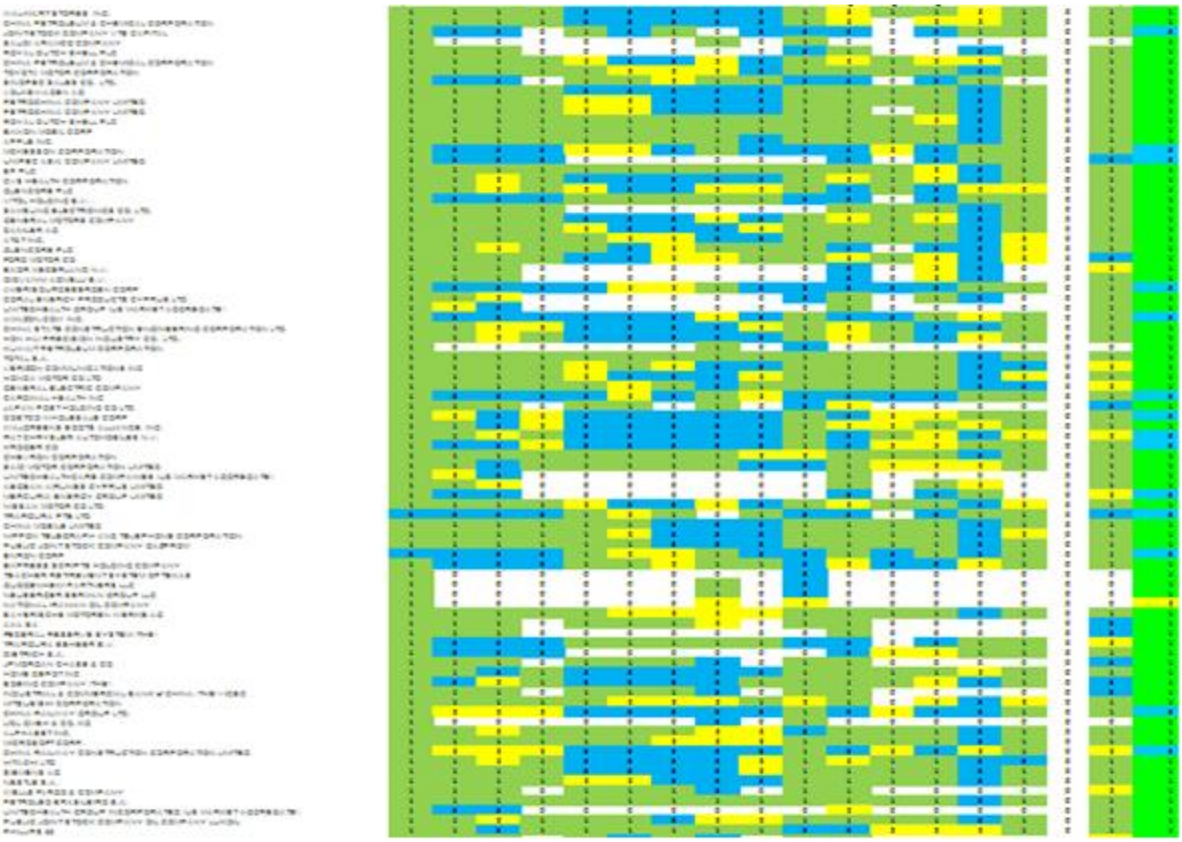

Figure 6: Result patterns of annual data of mixed companies generated by the consciousness model (excerpt).
Source: Own source. Own editing

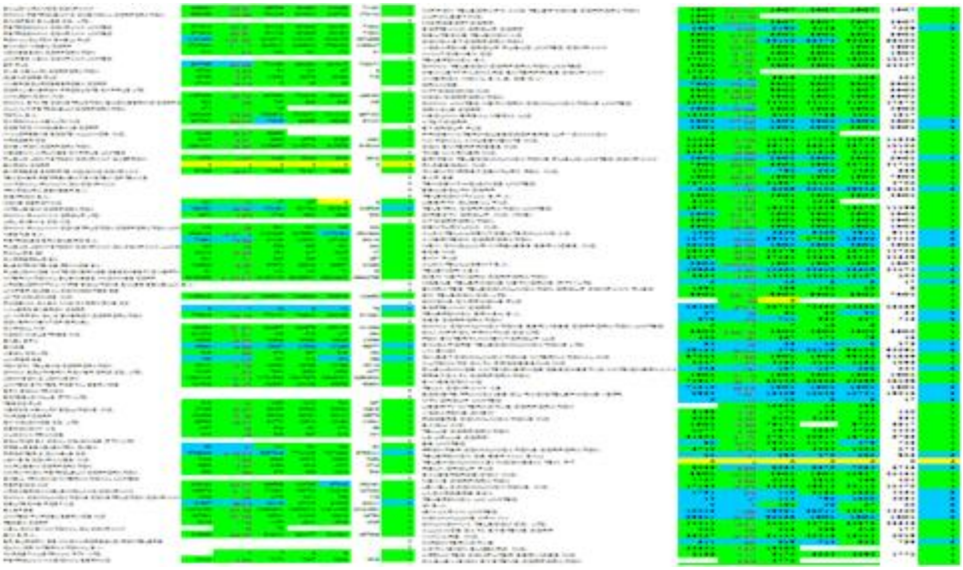

Figure 7: The innovation element's pattern in the annual data of the consciousness model of NAICS2012 and Infoalldt companies (excerpt)
Source: Own source. Own editing

Figure 8: The innovation element's pattern in the quarterly data of the consciousness model of NAICS2012 and Infoalldt companies (excerpt).
Source: Own source. Own editing

3. Applicability of the results and future research directions

My study was primarily directed at the application of neural networks assisting *non-deep learning*, for the recognition of models and the displaying of correlations. The even more accurate functioning of the model can be facilitated by the application of neural networks assisting *deep learning*, thereby according the basic principles they become adaptable for other independent data systems.

The type of consciousness model outlined by me uses the most basic components and their correlations. I constructed the model while considering the defining and significant elements. The system can be further refined by program elements applying artificial intelligence. The model and the selection system are applicable to all areas of economics, and can be further developed independent of the type of company groups.

References

[1] Alam, P.-Booth, D.,-Lee, K.-Thordarson, T. (2000) The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study. *Expert Systems with Applications.* 18 (3). pp.185-199.

[2] Ahn, B.S.-Cho, S.S.-Kim, C.Y. (2000) The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications* 18 (2). pp.65-74.

[3] Anandarajan, M.-Anandarajan, A. (1999) A comparison of machine learning techniques with a qualitative response model for auditor's going concern reporting. *Expert Systems with Applications*. 16 (4), pp.385-392.

[4] Back B.-Toivonen, J.-Vanharanta, H.-Visa, A. (2001) Comparing numerical data and text information from annual reports using self-organizing maps. *International Journal of Accounting Information Systems* 2 (4), pp. 249-269.

[5] Ballarin, A. S.-Gervasi-V. Cannat-S. Liudaki (1995) Company Financial Strategic Analisys Using Neural Classifiers. Computational Intelligence for Financial Engineering. IEEE.

[6] Berry, R. H.-S. Nix (1991) Regression Analysis v. Ratios in the Cross-section Analysis of Financial Statements. *Accounting and Business Research*. 21(82), pp. 107-117.

[7] Benhayoun, N.-Chairi, I.-El Gonnouni, A.-Lyhyaoui, A. (2013) Financial intelligence in prediction of firm's creditworthiness risk: evidence from support vector machine approach. *Procedia Economics and Finance* 5, pp. 103-112.

[8] Boyacioglu, MA-Kara, Y.-Baykan, O.K. (2009) Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: a comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications* 36 (2), pp. 3355-3366.

[9] Calderon, T.G.-Cheh, J.J.(2002) A roadmap for future neural networks research in auditing and risk assessment. *International Journal of Accounting Information Systems* 3 (4), pp. 203–236.

[10] Callen, J.L.-Kwan, C.C.-Yip, P.C.-Yuan, Y. (1996) Neural network forecasting of quarterly accounting earnings. *International Journal of Forecasting* 12(4), pp. 475-482.

[11] Chakraborty, S.-Sharma, S.K. (2007) Prediction of corporate financial health by artificial neural network. *International Journal of Electronic Finance* 1 (4), pp. 442-159.

[12] Cleonabula M. M. Neves-Vitor R. M. Silva-Gabriella A. B. Barros-Roberta V. V. Lopes (2011): A Variation of the Genetic Algorithm of Holland to Support Analysis of Balance Sheet and Income Statement for the Fiscal Year. *Information Technology and Artificial Intelligence.* 2. p. 331.

[13] Coakley, J.R.-Brown, C.E. (2000) Artificial neural networks in accounting and finance: modeling issues. *International Journal of Intelligent Systems in*

Accounting, Finance & Management 9 (2), pp. 119–144.

[14] Chen, Y.-Zhang, L.-Zhang, L. (2013) Financial distress prediction for Chinese listed manufacturing companies. *Procedia Computer Science* 17, pp. 678-686.

[15] Chi, L.C, Tang, T.C., (2005): Neural networks analysis in business failure prediction of Chinese importers: a between-countries approach. *Expert Systems with Applications*. 29 (2), pp. 244-255.

[16] Dattilo G.-Greco S.-Masciari E.-Pontieri L. (2000) A Hybrid Technique for Data Mining on Balance-Sheet Data In.: Yahiko Kambayashi Mukesh Mohania A Min Tjoa (Eds.) Data Warehousing and Knowledge Discovery Second International Conference. London. *Springer*. pp. 419-425.

[17] Doumpos, M.-Gaganis, C.-Pasiouras, F. (2005) Explaining qualifications in audit reports using a support vector machine methodology. *Intelligent Systems in Accounting, Finance and Management* 13 (4), pp. 197-215

[18] Eklund, T.-Back, B.-Vanharanta, H.-Visa, A. (2008) A face validation of a SOM-based financial benchmarking model. *Journal of Emerging Technologies in Accounting* 5 (1), pp. 109-127.

[19] Etheridge, H.L.-Sriram, R.S.-Hsu, H.Y. (2000) A comparison of selected artificial neural networks that help auditors evaluate client financial viability. *Decision Sciences* 31 (2),pp. 531-550.

[20] Evans, J.R. (2013):Business Analytics: Methods, Models, and Decisions. *Prentice-Hall*, Boston.

[21] Ezazi, M.E.-Ghotbi, F.S.-Ghotbi, S.F. (2013) Predicting earning management using RBF, ICA, and SVM in firms listed in Tehran security exchange. *Asian Journal of Management Research* 4 (1), pp. 208-220.

[22] Farzaneh A. Amania-Adam M. Fadlalla (2017) Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems* 24. pp. 32–58.

[23] Fayyad, U. Piatetsky-Shapiro, G.-Smyth, P. (1996) From data mining to knowledge discovery in databases. *AI Magazine*. 17 (3), p. 37.

[24] Foltin, C.-Garceau, L. (1996) Beyond expert systems: neural networks in accounting. *National Public Accountant* 41 (6), pp. 26-32.

[25] Gray, G.L.-Debreceny, R.S. (2014) A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information System*. 15 (4), pp. 357-380.

[26] Han, Jiawei-Micheline Kamber (2006) Data Mining: Concepts and Techniques, Second Edition. *Elsevier*. pp. 67-68.

[27] Hofmann, E.-Lampe, K. (2013) Financial statement analysis of logistics service providers: ways of enhancing performance. *International Journal of Physical Distribution and Logistics Management*. 43 (4), p.4.

[28] Hoglund, H., (2013b) Fuzzy linear regression-based detection of earnings management. *Expert Systems with Applications*. 40 (15), pp. 6166-6172.

[29] Hoglund, H. (2012) Detecting earnings management with neural networks. *Expert Systems with Applications*. 39 (10). pp. 9564-9570.

[30] Huang, S.M.-Tsai, C.F.-Yen, D.C.-Cheng, Y.L. (2008) A hybrid financial analysis model for business failure prediction. *Expert Systems with Applications*. 35 (3), pp. 1034-1040.

[31] Ittner, C.D.-Larcker, D.F. (2001) Assessing empirical research in managerial accounting: a value-based management perspective. *Journal of Accounting and Economics*. 32 (1), pp. 349-410.

[32] Kloptchenko, A.-Eklund, T.-Karlsson, J.-Back, B.-Vanharanta, H.-Visa, A. (2004) Combining data and text mining techniques for analyzing financial reports. *Intelligent Systems in Accounting, Finance and Management* 12 (1), pp. 29-41.

[33] Koh, H.C.-Tan, S.S. (1999) A neural network approach to the prediction of going concern status. *Accounting and Business Research*. 29 (3), pp. 211-216.

[34] Koh, H.C.-Low, C.K. (2004) Going concern prediction using data mining techniques. Managerial Auditing Journal. 19 (3), pp. 462-476.

[35] Koskivaara, E. (2004a) Artificial neural networks in analytical review procedures. *Managerial Auditing Journal*.19 (2), pp. 191-223.

[36] Lenard, M.J.-Alam, P.-Madey, G.R. (1995) The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. *Decission Sciences*. 26 (2), pp. 209-227.

[37] Leon, Carlos (2017) Whose Balance Sheet Is This? Neural Networks for Banks' Pattern Recognition.*Wilmott*. (91) pp. 34-47.

[38] Li, H.-Sun, J.-Li, J.C.-Yan, X.Y. (2013) Forecasting business failure using two-stage ensemble of multivariate discriminant analysis and logistic regression. *Expert Systems with Applications*. 30 (5), pp. 385-397.

[39] Magnusson, C.-Arppe, A.-Eklund, T.-Back, B.-Vanharanta, H.-Visa, A. (2005): The language of quarterly reports as an indicator of change in the company's financial status. *Information and Management*. 42 (4), pp. 561-574.

[40] Martens, D.-Bruynseels, L.-Baesens, B.-Willekens, M.-Vanthienen, J.(2008): Predicting going concern opinion with data mining. *Decision Support Systems*. 45 (4), 2008. pp. 765-777.

[41] Ngai, E.W.T.-Hu, Y., Wong-Y.H., Chen,Y.,-Sun, X., (2011) The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. *Decision Support Systems*. 50 (3), pp. 559–569.

[42] Peel, M.J.(1989) The going-concern qualification debate: some UK evidence. *British Accounting Review*. 21 (4), pp. 329-350.

[43] Pujari, A.K., (2001) Data Mining Techniques. *Universities press*.

[44] Quek, C.-Zhou, R.W.-Lee, C.H. (2009) A novel fuzzy neural approach to data reconstruction and failure prediction. *Intelligent Systems in Accounting, Finance and Management* 16 (1/2). pp. 165-187.

[45] Salterio, S. (1996) The effects of precedents and client position on auditors' financial accounting policy judgment. *Accounting, Organizations & Society* 21 (5). pp. 467-486.

[46] Shirata, C.Y.-Sakagami, M.(2008) An analysis of the Going Concern Assumption: text mining from Japanese financial reports. *Journal of Emerging Technologies in Accounting* 5 (1), pp. 1-16.
[47] Spathis, C.T. (2003) Audit qualification, firm litigation, and financial information: an empirical analysis in Greece. *International Journal of Auditing* 7(1), pp. 71-85.
[48] Song, D.B.,-Lee, H.Y.-Cho, E.J. (2013) The association between earnings management and asset misappropriation. *Managerial Auditing Journal*. 28 (6), pp. 542-567.
[49] Tang, T.C., Chi, L.C. (2005) Neural networks analysis in business failure prediction of Chinese importers: a between-countries approach. *Expert Systems with Applications*. 29 (2), pp. 244-255.
[50] Tsaih, R.H.-Lin, W.Y.-Huang, S.Y. (2009) Exploring fraudulent financial reporting with GHSOM. Proceeding of Pacific Asia Workshop (PAISI) *In Intelligence and Security Informatics.*
[51] Tung, W.L.-Quek, C.-Cheng, P. (2004) GenSo-EWS: a novel neural-fuzzy based early warning system for predicting bank failures. *Neural Network*. 17 (4), pp. 567-588.
[52] Zhang, Li-Lu Wang-Jianping Zhang (2011) A Computer Auditing Model of the Balance Sheet Parallel Simulation Based On Data Mining. In. Modeling Risk Management for Resources and Environment in China. *Springer-Verlag*. Berlin Heidelberg. pp. 567-576.
[53] Yang, J.G.S. (2006) Data mining techniques for auditing attest function and fraud detection. *Journal of Forensic Investigative Accounting* 1 (1), pp. 4–10.
[54] Yang, Steve, Randy Cogill (2013) Balance Sheet Outlier Detection using a Graph Similarity Algorithm. *2013 IEEE Conference on Computational Intelligence for Financial Engineering & Economics*. pp. 1-8.
[55] Yang, Jun – Zhang,Yuli – Au,Kevin – Xue,Hongzhi (2011) Prior experiences and social class as moderators om planning-performance relationship in China’s new business ventures. *National Science Foundation of China*. pp.1-40.
[56] Youn, H.-Gu, Z. (2010) Predicting Korean lodging firm failures: an artificial neural network model along with a logistic regression model. *International Journal Hospital Management.* 29 (1), pp. 120-127.
[57] Wang, S. (2010): A comprehensive survey of data mining-based accounting-fraud detection research. *Proceeding of the International IEEE Conference on Intelligent Computation Technology and Automation*.