Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 176 (2020) 3037–3046

Procedia Computer Science

www.elsevier.com/locate/procedia

24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

Application of machine learning models and artificial intelligence to analyze annual financial statements to identify companies with unfair corporate culture

Joanna Wyrobek[a]*

[a]Department of Corporate Finance, Cracow University of Economics, ul. Rakowicka 27, 31-510 Kraków, Poland

Abstract

The purpose of the publication was to create a model that, based on the annual financial statements, identifies the risk of significant financial irregularities occurring in the enterprise. These irregularities may relate to different types of financial fraud that do not necessarily affect the annual financial statements. A characteristic feature of irregularities is that they are large-scale and will have a drastic impact on the company's reputation. The results of the research show that machine learning and artificial intelligence algorithms were able to learn to recognize patterns of such scams and can detect them very effectively. An element of the novelty of the presented research is that it shows the possibility of training algorithms to recognize fraud based on information that is often not related directly to the observed fraudulent activities. The practical importance of research is the possibility of using the model in the decision-making process in the enterprise. The model allows assessing the risk that a potential business partner may commit financial fraud, which requires careful examination of the integrity of such an enterprise.

Keywords: fraud detection; machine learning; artificial intelligence;

* Corresponding author. Tel.: +48-793643705; fax: +48-12-2935081.
E-mail address: wyrobekj@uek.krakow.pl

10.1016/j.procs.2020.09.335

1. Introduction

Dishonesty in doing business is nothing new. Even though history teaches that such a policy does not pay off in the long run, there is still a lot of information in the press about financial scandals that have taken place in various countries. The industrial revolution 4.0 that is currently taking place, on the one hand, makes it easier to commit a lot of fraud, but on the other hand, it also provides useful new tools for identifying dishonest companies with fraudulent corporate culture.

The most famous fraud detection models are the m-score model (Beneish [2]) and the F-score model (Dechow et al.[4]). Beneish model was based on an unweighted probit, and Dechow's model was based on the logistic regression. Both methods require the calculation of several financial ratios, which are substituted for the linear combination of these ratios with fixed parameters. If the resulting value exceeds a certain threshold, then it is interpreted as a risk that the financial statement could have been manipulated.

Since the development of Beneish and Dechow's models, financial statements fraud became a prevalent research topic, and many models have been created which develop their idea. Most of these models identify which financial statement is manipulated based on its contents. Very few models use one annual financial report to predict that the next annual financial statement is going to be manipulated. Only a handful of the models use quarterly data. If they do, they mix annual reports with quarterly reports in one model, despite different valuation and classification rules which are used for annual and quarterly statements.

All the models which we found concentrated either solely on earnings manipulation or purposeful manipulation of financial statements, which became the subject of the official authorities investigation (mostly SEC – Securities and Exchange Commission). Models analyze financial statements looking for the patterns which identify purposeful manipulation of this financial statement.

The purpose of this paper was to verify whether it is possible to train an efficient model that uses annual financial statements (they are available not only for publicly traded companies but also for private businesses) to identify companies which in the same year committed a significant fraudulent transaction. This transaction could but did not have to affect this annual financial statement directly. The committed fraud could not have a direct impact on the financial statement of an enterprise. We believe, however, that fraudulent behavior could indirectly affect the report and that it created fraudulent patterns in this report, which can be identified with machine learning or artificial intelligence algorithms.

The research problem verified in this paper is whether the analysis of annual financial reports can warn about the risk of severe financial fraud going on in the company, even if it is not affecting this statement directly.

2. Literature review

Table 1 shows a summary of selected papers dedicated to financial fraud detection models. As can be seen, most researchers used annual financial statements and wanted the model to identify which ones were manipulated. In two cases, we found the application of both: annual and quarterly data in the same model. In one paper, the researchers attempted to create the model, which predicted that the next financial statement is going to be manipulated. The most famous and popular models shown in Table 1 are the Beneish and Dechow et al. models. One can find these models used on financial information websites as a part of the financial analysis of companies.

Table 1. Financial statement manipulation detection models

Data type	What model detects	Best algorithm accuracy/recall	Best algorithm	time	country	source of fraud info	no of fraud/no-fraud firms	authors
Financial ratios (FR)	fraudulent financial statements (FSS)	71,7/68,4	ANN	n.a.	USA	SEC detected fraudulent statements	46/46	Green, Choi[11]
FR	FSS	83,54/81,08	LR	1980-1990	USA		77/305	Bell, Carcello [1]

Data type	What model detects	Best algorithm accuracy/recall	Best algorithm	time	country	source of fraud info	no of fraud/no-fraud firms	authors
FR	FSS	71,5/64	LR	1970-1990	USA	SEC charged intentional material misstatement	100/100	Persons [21] (with 1 year advance)
FR+Corporate Governance	FSS	62,5/66	ANN	n.a.	USA	COMPUSTAT financial restatements, any litigation which was disclosed to SEC	102/102	Fanning, Cogger [7]
FR, insider trading factors	FSS	66,7/72,2	LR	1973-1988	USA	The Wall Street Journal Index for cases identified as financial statement fraud	51/51	Summers, Sweeney [25]
FR	targets of SEC investigations	69,72/81,03	ANN	n.a.	Greece	SEC detected fraudulent statements	38/38	Feroz et al. [8]
FR	FFS	87,75/86,29	UTADIS	n.a.	Greece	auditors opinion or investigation against financial statement fraud	38/38	Spathis et al. [24]
FR	FFS	76/35	FNN	1980-1995	USA	SEC charged for fraudulent statements	40/160	Lin et al. [20]
FR	FFS	53,8/21,7	LDA	1975-1999	USA	SEC enforcement against fraudulent statements	79/79	Kaminski et al. [15]
FR	earnings manipulation (EM)	67/81,3	Three-phase cutting plane	1992-2002	Turkey	earnings restatements	126/168	Dikmen, Küçükkocaoğlu [5]
FR	EM	89.5/54,2	unweighted profit	1982-1992	USA	SEC enforcement against fraudulent statements	50/1708	Beneish [2]
FR	EM	64,41/65,59	LR	1982-1993	USA (quarterly and annual)	SEC enforcement against fraudulent statements	451/130K	Dechow et al. [4]
FR	FFS	90,3/91,7	Bayesian Belief Network	n.a.	Greece	auditors opinion or investigation against financial statement fraud	38/38	Kirkos et al. [16]
FR	FFS	95/63	genetic algorithms	1991-2003	USA (annual and quarterly)	SEC charges improperly recognizing revenue	51/339	Hoogs et al. [13]
FR	FFS	65,8/71,3	SVM	1995-2004	USA	SEC enforcement against fraudulent statements	101/101	Humphreys et al. [14]
FR	FFS	38,46/45,16	Text mining	2000-2008	USA	SEC enforcement against fraudulent statements	11/20	Glancy [10]
FR	FFS	98,09/98,09	Probabilistic NN	n.a.	USA	not disclosed	101/101	Ravisakar [23]

Data type	What model detects	Best algorithm accuracy/recall	Best algorithm	time	country	source of fraud info	no of fraud/no-fraud firms	authors
FR	FFS	95,1/90,02	Stacking	2001-2002	Greece	auditors opinion or investigation against financial statement fraud	41/123	Kotsiantiset et al. [17]
Numerical financial variables, onthology	management fraud	90,4/80	SVM	1982-2005	USA	SEC enforcement against fraudulent statements	205/6427	Cecchini et al. [3]
Analyst calls+Financial variables	financial restatements	89,03/24,69	LR	2003-2007	USA	financial restatements	29663/29663 transcripts	Larcker and Zakolyukina [18]
Financial+non-financial variables	FFS	89,02/n.a.	UTADIS	2001-2004	Greece	auditors opinion or investigation against financial statement fraud	199/199	Gaganis [9]
Management report	FFS	82,95/80,71	LR	1994-2009	USA	SEC enforcement against fraudulent statements	1058/5534 reports	Purda and Skillicom [22]
Traits from financial social media+financial variables+management reports	corporate fraud	80/83,04	SVM	2004-2015	USA	SEC announcements about companies' fraud	64/64	Dong et al. [6]
FR+risk factors +organization risk + earnings stability measures + risk attitures	FFS	93,3/86,7	logit with the variables based on fuzzy-logic	1996-2001	USA	SEC enforcement against fraudulent statements	15/15	Lenard et al. [19]

If one looks at Table 1, it is visible that machine learning and artificial intelligence algorithms (neural networks) perform better than linear regression or logistic regression when they are applied to the same data. Still, the most popular models are a white-box type (as in opposition to black-box models) because they are easily transferable and can be easily used by everyone. Most models only use annual data because it is usually the manipulation of yearly data that is important for the shareholders. Authors use not only financial statements but also management reports, analyst calls, analysis of speech, and images for character traits as well as social media opinions, mainly taken from investor forums.

From the point of view of this research, the most adequate are the results of Dong et al. [6] and Fanning and Cogger[7] because these authors collected data not only regarding counterfeiting of financial statements but also other scams (based on the description of the method of selecting the research sample).

Similarly to these authors, we took the list of largest financial scandals of the XXth century and made a list of companies that not only manipulated their annual earnings, but also: manipulated their quarterly statements, escaped taxes, used special purpose vehicles to hide illegal transactions, used price collusion, bribery, corrupted other countries' authorities, and concealed their activities through complex offshore holding structures. We have found that even in cases where these activities were not associated directly with the falsification of the financial statements, the effects of these activities may still be visible in the financial report. To facilitate model learning, we used only the most significant financial scandals to have a better chance of teaching such a model successfully.

3. Research method description

To check the capability of the annual statements to identify corrupted and fraudulent corporate culture (and behavior), we collected financial statements of 54 companies which are enlisted on Wikipedia as a list of biggest accounting scandals of the XX-th century. We matched them with a sample of 58 similar "honest" companies in terms of industry, size, type of activities (the algorithms trained on 1317 financial statements). Most of the companies selected for the model were listed on the NYSE or NASDAQ stock markets. To determine the years where a certain company was committing fraud, we used SEC investigations and press releases about bribery, tax evasion, and other scams committed by these companies.

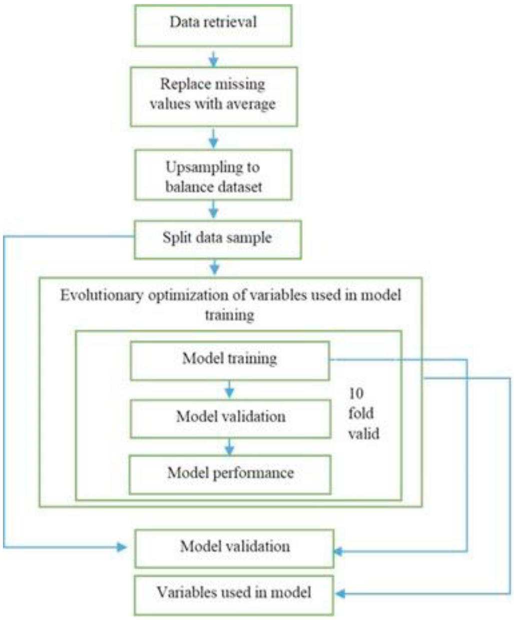

Figure 1. Training process structure

Data collected for the research was taken from the SEC commission website (which is freely available). We used the database of SEC investigations and the annual financial statements (10-k reports). For missing years and unlisted companies, we used the Thompson Reuters Worldscope database. We used all financial statements (balance sheet, income statement, cash flow statement) and financial ratios (current and quick liquidity ratio, acid test, ROA, ROE, EBT/Sales, GrossProfitOnSales/COGS, GrossProfitOnSales/Sales, NetProfit/Sales, Debt/Equity, LTCapital/FixedAssets, Liabilities/TotalAssets, NetWorkingCapital/TotalAssets, NetWorkingCapital/Sales, Days in Receivables, Days in Payables, Days in Inventories, InterestCover, Accurals/TotalAssets, Accurals/Sales, IntangibleAssets/TotalAssets, InstangibleAssets/FixedAssets, Provisions/TotalLiabilities, Provisions/TotalAssets, CFO/EBIT, CFO/Sales, CFO/(change of Fixed Assets + Dividends+changes of Debt), (Amortization+Depreciation)/FixedAssets), sales growth, fixed assets growth. We used for training 298 variables (every model had to select 15 of them). Elements of balance sheet statement were normalized by dividing them by total assets, the elements of the income statement were normalized by dividing them by total sales, and the elements of the cash flow were divided by sales receipts (which were estimated as total sales revenues minus a change in receivables).

Fig. 1 shows the process construction which we used for the model training. To balance the panel we used upsampling, which generated new members of the less numerous class, in this case, the unfair companies. Then the data sample was split in proportion 80/20 for training and validation. Then, the model was trained with the 10-fold validation method (on 80% of the data sample), and the genetic algorithm was using this method to select the best combination of variables.

For model training, we used the following algorithms: logistic regression, linear discriminant analysis, deep neural network, Naïve Bayes model, Support Vector Machines model, random forest model, and gradient boosted decision trees model. The models were created in Python with the help of the following libraries: Keras, TensorFlow, SciKit-learn NumPy, and pandas.

The general method of variable selection was based on a genetic algorithm, because the brute force method in the case of the need to select 15 variables out of 289 overloaded the computer's memory resources. Therefore, it was necessary to use a genetic algorithm to simplify the choice. The algorithm tested the number of variables for the model from 1 to 15. There were 5 representatives in each generation, and the maximum number of generations was 30. Further increasing the number of generations no longer improved the quality of the model. We tested different selection schemes (cut, uniform, Boltzmann, tournament, roulette wheel, stochastic universal sampling), and the best selection scheme turned out to be a tournament. The selection of variables for individually tested combinations was based on a set probability of 50%, also the cross-over probability was set at 50%.

In the case of logistic regression (LR), the only important setting was the standardization of variables and the use of the constant expression. Variables with a correlation above 40% we eliminated from the model to avoid collinearity.

For gradient boosted decision trees, we used the following settings: 100 decision trees with a maximum depth of 10 levels, the learning rate was set to 10% (and 20 bins).

For random forest training, we used the following parameters: 100 decision trees constructed with the gain ratio as a split criterion, and the maximum depth of 10 levels. As a voting strategy, we used the confidence vote. The algorithm guessed the subset ratio.

Deep Neural Network was tested for 1, 2, and 3 hidden layers and the number of neurons between 1 to 300 (due to a large number of observations). We chose the model with the highest average from 3 measures: accuracy, precision, and recall. We tested different activation functions (with and without a dropout), and we chose the tanh activation function. DNN used 10 epochs. We applied standardization. Rho parameter was set to 0,99 and epsilon to 1.0E-8.

For the Naïve Bayes model, we used Laplace correction to avoid false results if a given attribute never occurs in a given class (it was a default setting).

Linear Discriminant analysis did not involve any settings except for the removal of insignificant variables from the model (data was standardized).

Finally, the support vector machines model (SVM) was trained for the dot kernel type, the convergence epsilon 0,001, and the maximum number of iterations equal to 100 000. We used a balanced panel, but we turned on the option to balance the cost function.

4. Results and conclusions

4.1. Results

Table 2 shows the results of algorithms training. As can be seen in the table, the efficiency of the best algorithms was close to 95% on the separate evaluation sample (algorithms were trained on the training sample, and the presented results show their efficiency when they are applied to an independent, evaluation sample). On the one hand, we used only the most significant accounting scandals of the XXth century, so it would be surprising if the model could not detect them. On the other hand, many of these scams did not affect the annual statement of companies directly.

The most successful algorithms turned out to be the gradient-boosted decision trees (XGB) approach and the random-forest algorithm (R.F.). Gradient-boosted decision trees algorithm had a total accuracy of 93.5%, precision of 67,3%, and recall 76,1%. The random forest algorithm had an overall efficiency of 94,7%, precision of 80%, and recall of 76,1%. One of the "traditional" models, namely, linear discriminant analysis algorithm, also turned to be quite efficient with the overall accuracy of 92,2%, precision o 68,3%, and recall of 68,7%.

Table 2. Summary of own models identifying financial fraud

item	logit	XGB	RF	DNN	Bayes	LDA	SVM
accuracy	84,70%	93,50%	94,70%	91,04%	62%	92,20%	89,70%
precision	42,10%	67,30%	80%	72,39%	20,50%	68,30%	75,50%
recall	63,80%	92%	76,10%	81,06%	71,80%	68,70%	24,50%
TN	1011	1081	1123	1081	700	1102	1141
TN%	87,60%	93,67%	97,30%	93,67%	60,70%	95,50%	98,90%
FP	143	73	31	73	454	52	13
FP% (I-st type error)	12,40%	6,30%	2,68%	6,33%	39,30%	4,50%	1,10%
FN	59	13	39	45	46	51	123
FN% (I-nd type error)	36,20%	7,98%	23,90%	27,6%	28,20%	31,30%	75,40%
TP	104	150	124	118	117	112	40
TP%	63,80%	92%	76,10%	72,39%	71,80%	68,70%	24,50%

If we look at the confusion matrix parameters, for the gradient boosted decision trees algorithm, its first type error was 6,3%, and the second type error was 7,98%, which means that the algorithm classified honest companies as dishonest only in 7 cases out of 100. The algorithm classified fraudulent firms to be reliable businesses in 8 cases out of 100 cases. For the random forest algorithm, the first type error was 2,68%, and the second-type error was 23,9%. For the LDA, the first-type error was 4,5%, and the second-type error was 31,3%.

If we compare our results to previous models (especially Dong et al.[6] and Faninng and Cogger[5]), no research from table 1 concluded that gradient boosted trees algorithm was the best method to train the model. Most cited papers stated that the authors achieved the best model efficiency for logistic regression or logit models. Accuracy and recall differ significantly between researches, and they depend on the size of the data sample and the scale of irregularities that were included in the model. Therefore, it is not surprising that our model performed very well because it was trained on the most infamous cases of management misbehavior. Our model has a broader scope, but it also covers purposeful financial statement fraud.

Table 3 shows which variables (out of 289 variables used for the model training) were selected by different algorithms (random forest, LDA, etc.).

Table 2. Variables left in the models after dimensionality reduction

LR	GBX	Random Forest	DNN	NaiveBayes	LDA	SVM
CashfromFinancingActivities	AmortizationofIntangiblesCF	AmortizationofIntangiblesCF	AccumulatedDepreciationTotal	CashfromFinancingActivities	AccountsReceivableTradeNet	AmortizationofIntangiblesCF
DiscontinuedOperationsLiabilities	CashfromFinancingActivities	CashfromFinancingActivities	CashFromFinancingActivities	CashTaxesPaid	CashfromFinancingActivities	CashfromFinancingActivities
ExciseTaxesPayments	DiscontinuedOperationsLiabilities	CashfromOperatingActivities	DaysPayables	FuelPurchasedforResale	DiscontinuedOperationsLiabilities	DiscontinuedOperationsLiabilities
InterestExpenseNetNonOperating	ExtraordinaryItem	DebtToEquityRatio	ExtraordinaryItemCF	InterestExpenseNetNonOperating	ExciseTaxesPayments	ExtraordinaryItem
InterestIncomeNonBank	InterestExpenseNetNonOperating	ExtraordinaryItem	InterestOrInvestIncomeNonOperating	MinimumPensionLiabilityAdjustment	InterestExpenseNetNonOperating	InterestExpenseNetNonOperating
InterestOrInvestmentIncomeOperating	NetWorkingCapitalToSales_Ratio	ExtraordinaryItemCF	LTInvestmentAffiliateCompanies	MinorityInterest(BalanceSheetLiabilities)	InterestIncomeNonBank	NetWorkingCapitalToSales_Ratio
MinimumPensionLiabilityAdjustment	MinorityInterest(IncomeStatement)	FuelInventory	NetIncomeAfterTaxes	MinorityInterest(IncomeStatement)	InterestOrInvestmentIncomeOperating	MinorityInterest(IncomeStatement)
OtherEquityTotal	OtherCurrentLiabilities	NetWorkintCapitalToSales_Ratio	OtherCurrentLiabilities	OtherEquity	MinimumPensionLiabilityAdjustment	OtherCurrentLiabilities
OtherLongTermAssets	OtherEquityTotal	OtherCurrentLiabilities	OtherLiabilitiesCF	PensionBenefitsUnderfunded	OtherEquityTotal	OtherLongTermAssets
RestrictedCashCurrent	OtherLongTermAssets	OtherLongTermAssets	Provisions_TotalLiab	ProvisionforDoubtfulAccounts	OtherLongTermAssets	ReportedCashfromOperatingActivities
ShortTermInvestments	ReportedCashfromOperatingActivities	ReportedCashfromFinancingActivities	RedeemableConvertiblePreferredStock	RB_CostOfRevenue	RestrictedCashCurrent	RestrictedCashCurrent
TotalReceivablesNet	RestrictedCashCurrent	RestrictedCashCurrent	SaleofIntangibleAssets	SellingGeneralAdminExpensesTotal	ShortTermInvestments	ShortTermInvestments
UnrealizedGainLoss	ShortTermInvestments	ShortTermInvestments	ShortTermInvestments	TotalCurrentLiabilities	TotalReceivablesNet	TotalReceivablesNet

LR	GBX	Random Forest	DNN	NaiveBayes	LDA	SVM
UtilityPlantNet	TotalReceivablesNet	TotalReceivablesNet	TotalAssets	TotalReceivablesNet	UnrealizedGainLoss	WPP
	WPP	WPP	TotalReceivablesNet	UnrealizedGainLoss	UtilityPlantNet	

As can be seen in table 3, the LR algorithm selected only 14. For unfair companies, the following variables were higher: cash flows from financing activities, discontinued operations, investment income, interest income, other equity, short-term investments, unrealized gains, and losses. Unfair companies had the following variables lower than "honest" companies: excise taxes (income statement), interest payments, other assets, restricted cash, pension liabilities adjustments, total receivables, and plant assets. This suggests a relatively strong concentration on financial operations by unfair companies and less value invested in plant assets.

For gradient boosted decision trees the algorithm selected the following variables and they were higher for unfair companies (compared to fair companies): cash flows from financing activities, discontinued operations, extraordinary items (in the income statement), net working capital/sales, minority interest, other equity, reported cash from operating activities, short-term investments. Unfair companies had the following variables lower: accumulated amortization of intangibles, interest payments, other current liabilities, other long-term assets, restricted cash, total receivables, fast liquidity ratio. This suggests lower liquidity in a narrow sense, extensive financial operations, fewer investments in intangible assets, more discontinued operations.

Since many variables were repeated in the random forest algorithm compared to previously discussed algorithms, we will only discuss the new variables. According to the random forest algorithm, the new variables (not seen in previous models), which were important in identifying unfair companies and were higher for fraudulent firms included: debt to equity ratio. Unfair companies also had more negative extraordinary items in cash flows statement and lower (if any) fuel inventories. Higher indebtedness may result in higher bankruptcy risk, which could force the management to unethical behavior to save the company.

The deep neural network introduced the following new variables, which the algorithm selected to be significant: longer days payable for unfair firms, lower investments in affiliate companies, lower net income after taxes, higher provisions related to total liabilities, less convertible stock, more sales of tangible assets, lower total assets.

Naïve Bayes model (compare with table 2) introduced the following new variables: lower (for unfair companies) cash taxes paid, more fuel repurchased for resale (if any), the higher gross profit margin on sales, lower selling and administrative expenses, lower total current liabilities, higher unrealized gains /losses. This can be interpreted as higher profitability on sales, but more costs that reduce the profits (compared to trustworthy companies), fewer tax payments, and more transactions, which result in unrealized gains or losses at the end of the year.

The LDA model selected variables that were all discussed for previous models.

The SVM model selected variables that were already discussed for other models (it did not use any new variables from the dataset).

If we compare variables that our models found important in fraud detection with previous models (presented in table 1), there are some interesting observations. None of the algorithms applied in this study selected accruals as an important variable, and accruals are considered to be significant in most of the publications presented in table 1 (we did not estimate excessive accruals though, which is the most popular variable in FFS research because such a model depends on yet another regression which has to be estimated). Only the random forest algorithm used selected another variable popular in FFS models, which is the debt to equity ratio (as a measure of leverage). If we compare variables selected by the algorithms in this study to the Beneish model [2], both studies believe in the importance of the size of receivables, gross profit margin, size of depreciation and amortization, sales and general expenses, and leverage. Beneish model adds to this list: total accruals divided by total assets and the asset quality index. Dechow's model[4] is more concentrated on the growth of different elements of assets and profits (and accruals). Compared to Summers and Sweeney's model[25] the algorithms used in this study did not select popular profitability ratios such as the return on assets (Summers and Sweeney, however, also underlined the importance of inventories and receivables in fraud prediction which agrees with the results of this study).

Compared to previous models, the algorithms trained in this study showed the importance of the scale of financial operations, which included investing in financial assets, and transactions reported in the financial part of the cash flow statement. Algorithms also suggested the importance of discontinued operations (which are higher for unfair companies) and the cash payments for taxes. Unfair companies have a higher gross profit margin but lower net profits. Unfair companies also had lower interest payments. Unfair companies also had lower liquidity. All these variables suggest excessive concentration on the financial activities and obtaining capital from the market and low control over own costs and own operating activities.

4.2.Conclusions

The purpose of the paper was to test whether machine learning and AI algorithms can find patterns in annual financial statements that indicate a fraudulent corporate culture and committing various sorts of massive financial crime in a company. All such acts require an official investigation of the supervisory bodies, and our models identified such situations with the accuracy of close to 95%. So the first conclusion is that the evidence shows that such a model can be created, and it works well.

Obtained results suggest that unfair companies have a high gross profit margin, but low net profits and pay relatively lower taxes compared to honest firms. Dishonest companies also seemed to be more active in financial operations, which included obtaining new capital and making investments in financial assets. Unfair companies also tended to have relatively lower liquidity ratios, more extraordinary items and discontinued operations, and higher indebtedness.

What is also worth mentioning, in the paper we presented an approach where algorithms in which training results are difficult to interpret (such as ensemble models and deep neural network) were combined with genetic algorithms to derive a set of most significant variables which can be interpreted.

Acknowledgments

The publication is financed by the Ministry of Science and Higher Education of Poland within ―Regional Initiative of Excellence‖ Programme for 2019-2022. Project no.: 021/RID/2018/19. Total financing: 11 897 131,40 PLN.

References

[1] Bell, Timothy B., and Joseph V. Carcello. (2000) "A decision aid for assessing the likelihood of fraudulent financial reporting." *Auditing* **19 (4)**: 168–184.

[2] Beneish, Messod D. (1999) "The Detection of Earnings Manipulation." Financial Analysis Journal **55 (5)**: 24–36.

[3] Cecchini, Mark, Haldun Aytug, Gary J. Koehler, and Praveen Pathak. (2010) "Detecting management fraud in public companies." *Management Science* **56 (7)**: 1146–1160.

[4] Dechow, Patricia M., Weili Ge, Chad R. Larson, and Richard G. Sloan. (2011). "Predicting Material Accounting Misstatements." *Contemporary Accounting Research* **28 (1)**: 17–82.

[5] Dikmen, Burcu, and Guray Ķ̈kkocaǒlu. (2010) "The detection of earnings manipulation: The three-phase cutting plane algorithm using mathematical programming." *Journal of Forecasting* **29 (5)**: 442–466.

[6] Dong, Wei, Shaoyi Liao, and Zhongju Zhang. (2018) "Leveraging Financial Social Media Data for Corporate Fraud Detection." *Journal of Management Information Systems* **35 (2)**: 461–487.

[7] Fanning, Kurt M., and Kenneth O. Cogger. (1998) "Neural network detection of management fraud using published financial data." *Intelligent Systems in Accounting, Finance, and Management: An International Journal* **7 (1)**: 21–41.

[8] Feroz, Ehsan H., Taek Mu Kwon, Victor Pastena, and Kyung Joo Park. (2000) "The efficacy of red flags in predicting the SEC's targets: an artificial neural networks approach." *Intelligent Systems in Accounting Finance & Management* **29 (3)**: 145–157.

[9] Gaganis, Chrysovalantis. (2009) "Classification techniques for the identification of falsified financial statements: a comparative analysis." *Intelligent Systems in Accounting, Finance & Management* **16 (3)**: 207–229.

[10] Glancy, Fletcher H., and Surya B. Yadav. (2011) "A computational model for financial reporting fraud detection." *Decision Support Systems* **50 (3)**: 595–601.

[11] Green, Bryan Patrick, and Choi J. (1997) "Assessing the risk of management fraud through neural network technology." *Auditing* **16 (1)**: 25–28.

[13] Hoogs, Bethany, Thomas Kiehl, Christina Lacomb, and Deniz Senturk. (2007) "A Genetic Algorithm Approach to Detecting Temporal Patterns Indicative of Financial Statement Fraud." *Intelligent Systems in Accounting, Finance & Management* **15 (1-2)**: 41–56.

[14] Humpherys, Sean L., Kevin C. Moffitt, Mary B. Burns, Judee K. Burgoon, and William F. Felix, "Identification of fraudulent financial statements using linguistic credibility analysis," *Decision Support Systems* **50 (3)**: 585–594.

[15] Kaminski, Kathleen A. T. Sterling Wetzel, and Liming Guan, "Can financial ratios detect fraudulent financial reporting?" *Managerial Auditing Journal* **19 (1)**: 15–28.

[16] Kirkos, Efstathios, Charalambos Spathis, and Yannis Manolopoulos. (2007) "Data Mining techniques for the detection of fraudulent financial statements." *Expert Systems with Applications* **32 (4)**: 995–1003.

[17] Kotsiantis, Sotiris, E. Koumanakos, D. Tzelepis, and V. Tampakas. (2006) "Forecasting fraudulent financial statements using data mining." *International Journal of Computational Intelligence* **3 (2)**: 104–110.

[18] Larcker, David F., and Anastasia A. Zakolyukina. (2012) "Detecting Deceptive Discussions in Conference Calls." *Journal of Accounting Research* **50 (2)**: 495–540.

[19] Lenard, Mary Jane, Ann L. Watkins, and Pervaiz Alam. (2007) "Effective Use of Integrated Decision Making: An Advanced Technology Model for Evaluating Fraud in Service-Based Computer and Technology Firms." *Journal of Emerging Technologies in Accounting* **4 (1)**: 123–137.

[20] Lin, Jerry W., Mark Hwang, and Jack D. Becker. (2003) "A fuzzy neural network for assessing the risk of fraudulent financial reporting." *Managerial Auditing Journal* **18 (8)**: 657–665.

[21] Persons, Obeua. (1995) "Using financial statement data to identify factors associated with fraudulent financial reporting." *Journal of Applied Business Research* **11 (3)**: 38–46.

[22] Purda, Lynette, and David Skillicorn. (2015) "Accounting Variables, Deception, and a Bag of Words: Assessing the Tools of Fraud Detection." *Contemporary Accounting Research* **32 (3)**: 1193–1223.

[23] Ravisankar, Pediredla, Vadlamani Ravi, Gundumalla Raghava Rao, and Indranil Bose. (2011) "Detection of financial statement fraud and feature selection using data mining techniques." *Decision Support Systems* **50 (2)**: 491–500.

[24] Spathis Charalambos, Michalis Doumpos, and Constantin Zopounidis. (2002) "Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques." *European Accounting Review* **11 (3)**: 509–535.

[25] Summers, Scott L., and John T. Sweeney. (1998) "Fraudulently misstated financial statements and insider trading: An empirical analysis." *Accounting Review* **73 (1)**: 131–146.