# **Accounting and Finance Review**

Journal homepage: www.gatrenterprise.com/GATRJournals/index.html



Acc. Fin. Review 2 (1) 1 – 8 (2017)

# The Comparison of Two Data Mining Method to Detect Financial Fraud in Indonesia

# Tarjo<sup>1</sup>\* and Nurul Herawati<sup>2</sup>

<sup>1,2</sup> University of Trunojoyo Madura, Raya Telang Street PO BOX 2, Bangkalan Madura 69192, Indonesia

#### ABSTRACT

GATR JOURNALS

**Objective** – This research is expected to improve the weaknesses in the research conducted by Tarjo and Herawati (2015). The objective of this study was to analyse two data mining methods in detecting financial fraud based on Beneish m-score model.

**Methodology/Technique** – The research data were companies who committed fraud based Database Case Sanctions Issuers and Public Companies which was released by the Financial Services Authority in the period 2001-2014. For comparison, researchers also used data from companies that did not commit fraud. Companies were selected based on the same industry group of companies committing fraud for the purposes of classification.

**Findings** – The results show that data mining methods can be used to detect financial fraud based on Beneish m-score model. However, there are differences in the classification. In the logit regression, the results are only limited to the accuracy of classification and weak. While the K-Nearest Neighbor model, in addition, it is capable of performing high classification accuracy.

**Novelty** – The study indicates a better method for detecting financial fraud. **Type of Paper:** Empirical

*Keywords*: Detecting Financial Fraud, Beneish M-Score, Logit Regression, K-Nearest Neighbor. JEL Classification: C81, M41.

#### **1. Introduction**

Study on detecting financial fraud has been conducted by some researchers. In the last years, some researchers used Beneish m-score model (Beneish, 1999; Beneish, Lee, & Nichols, 2012; Dimitriois, 2014), and data mining (Zaki & Theodoulidis, 2013; Ata and Seyrek, 2009). Meanwhile, study result of Tarjo and Herawati (2015) shows that study data indicates that classification of a company that conducts financial fraud is by 77.1%, while classification of a company that does not conduct financial fraud is by 80%.

The previous study is considered to be weak in term of accuracy in conducting detection. Therefore, this follows up study is expected to be able to increase classification accuracy in detecting company financial fraud. In methodology, analysis tool which is believed to be able to increase classification capability in

\* Corresponding author:

E-mail: tarjo@trunojoyo.ac.id

<sup>&</sup>lt;sup>+</sup> Paper Info: Received: January 2, 2017

Accepted: February 28, 2017

Affiliation: University of Trunojoyo Madura, Indonesia

detecting financial fraud is by using K-Nearest Neighbor (K-NN) analysis. As stated by Prasetyo (2012a, 2014) K-Nearest Neighbor (K-NN) method uses several K nearest neighbors on every class of a test data. This is expected to be an additional value and be more reliable in conducting classification of company financial fraud detection.

The main contribution of this study, theoretically, is about detection concept of company financial fraud and provides understanding on characteristics of the financial statement of a company that is indicated to commit financial fraud. This helps to describe clear empirical consequence of financial statement fraud. This study is a clear reference on detecting company financial fraud as the previous study conducted by Tarjo and Herawati (2015), Gaganis (2009), Kotsiantis, Koumanakos, Tzelepis, and Tampakas (2006). The purpose of this study was to analyze two data mining methods in detecting financial fraud based models Beneish mscore.

#### 2. Detecting Financial Fraud

Financial fraud has become an international phenomenon which occurs in most countries. As stated by Crumbley, Heitger, and Smith (2011), there are three types of financial fraud in financial statement which are first, manipulation, fraud, or change in accounting records or supporting documents of the prepared financial statement. Second, intentional errors or default of financial statement whether on the event, transaction, or other crucial information. Third, intentional misuse of accounting principles related to amount, classification, presenting procedures, or disclosure.

Financial statement fraud which is committed by management decreases company reputation in public view. Several financial fraud scandals have shown any kinds of its consequences. Based on the Association of Certified Fraud Examiners [ACFE] (2014), financial statement fraud is a fraud which is intentionally committed by manager or employees who do not state the actual financial statement information, for instance, fictive income, burden reported is too low. In the context of financial reporting, an auditory is designed to provide assurance to the concerned party that financial statement which is prepared by the company is not affected by material misstatement and provides adequate assurance against management accountability upon company asset as well (Koroy, 2008).

To maintain the consistency of this study, the researchers develop study roadmap. The result of a study conducted by Tarjo (2008, 2010), Tarjo and Herawati (2015) shows that go public company in Indonesia, especially for industrial is proved to commit profit modification/ profit management. Profit modification becomes the initial sign that fraud occurs in the company.

Findings of the study conducted by Omoye and Eragbhe (2014) states that investment and liquidity ratio is significantly related to financial statement fraud. This explains that financial ratio is worth criticized because it is possible that financial fraud indication exists. Meanwhile, the study conducted by Lenard and Alam (2009) explains that combination of accounting rules, regulations, and regulation enforcement have contributed to the development of fraud detection model. It is clear based on the study by Lenard and Alam (2009) that when fraud variable is added to bankruptcy prediction model (Beaver, McNichols, and Rhie, 2005; Ohlson, 1980, and Altman, 1968), bankruptcy model performance is increased. Furthermore, the model developed by Persons (1995) presents better fraud detection model with fraud variable addition and its capability to provide qualitative information. However, fraud risk factor evaluation consisting of fraud variable should also be continuously tested to reflect explicit indicators of fraud alerts availability. Therefore, this study tries to conduct classification study in performing financial fraud detection.

Findings by Sharma and Panigrahi (2012) indicates that data mining technique like a logistic model, neural network, Bayesian belief network, and decision trees have been applied the widest to provide the main solution against problems clinging in data fraud detecting and classifying. In another hand, a study by Chintalapati, Sowjanya, and Jyotsna (2013) has shown some data mining techniques to detect and prevent fraud with application in credit card and telecommunication that can be used to achieve higher cost saving and also for interest in determining potential legal evidence.

Asare, Wright, and Zimbelman (2015) that fraud investigator identifies the most important audit process for auditor in detecting fraud. Kirkos, Spathis, and Manolopoulos (2007) state that data mining technique has classification and prediction capabilities which can facilitate process decision making for financial problem. Financial and prediction tasks in literature which are collected discuss the topic of bankruptcy prediction, credit risk estimation, going concern, financial difficulties, company performance prediction, and management fraud.

The result of data mining application based on Terzi and Şen (2012) indicates that Artificial Neural Network (ANN) model is more effective compared to decision trees model. This study explains that data mining has an important role in assisting in making a decision. One of the decisions which are based on using data mining is to detect financial fraud availability.

A study by Ujal, Amit, Hiral, and Rajen (2012) detects fraudulent financial statement that may help auditor, tax authority and banker to identify false financial statement. This kind of study is required for Directorate General of Taxation to detect the availability of company financial fraud to avoid taxation. The study result by Suyanto (2009) indicates that pressure which is represented by net profit/ total assets and opportunities which are represented by supply/ total assets, other party transaction, and Big 4 (BIG4) of Audit Company correlate significantly to financial reporting fraud.

A study by Chongsirithitisak (2015) indicates that characteristics of creative accounting company consistent with the model. Financial characteristics are smaller in term of total assets, liquidity, and leverage of their partners. Those are a smaller size, selling, market value, and working capital against total assets, current ratio, and total debt to a total asset of their partners. In addition, they are less profitable and more leverage as well. In the other hand, selling growth is bigger than a company which is not creative accounting in the financial statement. As for capital index, the result indicates that there are four indexes that can show creative accounting in the financial statement, Days Sales in Receivable Index, Gross Margin Index, Assets Quality Index and Selling Growth Index. There is only one index that can show creative accounting in financial statement which is accrual against the number of assets.

Study result of Amara, Amar, and Jarboui (2013) show that performing culture given to the leadership is big pressure for fraud detection. Indeed, company stability, good imagery at workforce market, reputation and strive to increase visibility in the market are pressure evidence related to performance factors that cause the leader to commit fraud in the financial statement.

#### 3. Research Method

#### 3.1. Data

Data of this study were financial statements of companies which were indicated to commit fraud based on a release issued by Capital Market Supervisory Agency or Financial Services Authority in the period of 2001-2014. For classification purpose, then financial statements of companies that did not commit fraud in the same period of a year and the industrial group were used. Classification of data mining examination results in 2 (two) data groups which were training data and tested data. Training data consisted of 20 (twenty) companies which were indicated to commit fraud and companies who were not indicated to commit fraud and companies which were indicated to commit fraud and companies which were indicated to commit fraud and companies which were indicated to commit fraud.

Testing using data mining required attribute. This study used attribute based on size or variable in Beneish M-Score model (Beneish et al., 2012).

#### 3.2. Data Mining Testing

As study conducted by Tarjo and Herawati (2015), this study added data mining testing by using K-Nearest Neighbor method. This testing was conducted as comparison as well as to cover the weakness of the previous testing method. As for procedures of the data mining testing technique in this study were as follows:

- 1. Logistic Regression.
- 2. K-Nearest Neighbor.

The following was model of Logit Regression:

FRAUD =  $\beta 0+\beta 1DSRI + \beta 2GMI + \beta 3AQI + \beta 4SGI + \beta 5DEPI + \beta 6SGAI + \beta 7TATA + \beta 8LVGI + \epsilon i$ Description: FRAUD = dumy variable (1 for company that commits fraud and 0 for the contrary) DSRI= Days Sales in Receivable Index GMI = Gross Margin Index AQI = Asset Quality Index SGI = Selling Growth Index DEPI = Depreciation Index SGAI = Selling, General and Administrative Expenses Index TATA = Total Accrual to Total Assets LVGI = Leverage Index  $\epsilon i$  = Residual value

#### 3.3. Data mining using K-Nearest Neighbor

Based on Prasetyo (2012a), K-Nearest Neighbor (K-NN) method used some K Nearest neighbors on every class of tested data. K-NN framework used K-NN as the basis of the framework, in which tested data has a value of companies that committed fraud and those that did not commit fraud on every class in the interval [0.1]. The following was the company membership value that committed fraud.

$$\sum_{k}^{c} \mu_{ik} = 1, 0 \leq \mu_{ik} \leq 1$$

 $\mu_{ik} \text{was}$  company that committed fraud of tested data xi to class – k

Every tested data xi should be looked for K nearest neighbor on every class of company that committed fraud or company that did not commit fraud by using formula:

$$d(x_1, x_2) = \left(\sum_{j=1}^{p} |x_{ij} - x_{ij}| s\right) \frac{1}{s}$$

To determine class of prediction result of tested data xi, class with the highest company group membership value of company data that commit fraud. Formula used was as follows.

$$y' = \sum_{k=1}^{c} (u_{ik})$$

With:

Y' = class of prediction

c = the number of class

Description:

1 was company that committed fraud

2 was company that did not commit fraud



Figure 1. Nearest Neighbor Data Mining Testing Model

## 4. Findings and Discussion

The initial step performed in data mining testing on this study is testing data mining with logistic regression on data training group. The testing result of logistic regression is shown in Table 1 below.

		Predicted			
	Non Fraud	Fraud	Percentage Correct		
Non Fraud	16	4	80		
Fraud	6	14	70		
Overall Percentage			75		

Table 1. Logistic Regression	: Classification Table
------------------------------	------------------------

Based on Logistic Regression testing, it indicates that classification accuracy in detecting company with fraud indication is by 70% and the company which is not indicated to commit fraud is by 80%. Overall, logistic regression model on tested data has detecting capability by 75%. Meanwhile, data mining testing result with logistic regression on tested data group indicates as follows (Table 2).Data mining testing result using K-Nearest Neighbor shows the more robust result in performing classification as shown in Table 3.

#### Tarjo and Nurul Herawati

Row No.	KESIMPUL.	confidence(	.confidencei.	prediction(K.	DSRI	GM	AQI	SGI	DEPI	SGA	LEVI	ACCRUAL
1	1.0	0.500	0.500	1.0	0.162	0.862	1.101	1.696	0	0.559	0.916	-0.006
2	1.0	0.800	0.200	1.0	0.984	2.600	1.006	0.773	1.084	1.425	1.005	-0.009
3	1.0	0.500	0.500	1.0	0.841	1.154	2.461	1.075	0.989	0.847	1.012	0.111
4	1.0	0.800	0.200	1.0	0.958	1	1.004	1.051	-133.217	1.121	1.343	0.181
5	1.0	0.200	0.800	00	0.658	1.171	0.979	1.748	1.474	0.633	1.009	-0.155
6	1.0	0.400	0.600	0.0	0.751	1.386	1.458	0.894	1.001	1.050	0.793	-0.224
7	1.0	0.700	0.300	1.0	0.185	0.789	0.125	2.256	-2.864	0	0.348	0.036
8	1.0	0.700	0.300	1.0	1.039	1.195	0.580	2.055	-5.185	0.805	0.641	0.073
9	0.0	0.300	0.700	0.0	0.999	1.513	1.102	1.138	1.027	1.224	1.128	0.054
10	0.0	0.400	0.600	0.0	1.201	0.836	3.064	1.350	1.163	1.094	0.805	-0.009
11	0.0	0.200	0.800	0.0	1.036	1.086	0.731	1.113	1.020	1.108	0.924	0.101
12	0.0	0.300	0.700	0.0	0.671	1.471	0.190	0.935	0.744	1.367	0.887	-0.185
13	0.0	0.200	0.800	0.0	0.895	0.910	0.519	1.252	0.922	1.036	0.830	-0.113
14	0.0	0.200	0.800	0.0	0.863	1.010	1.352	1.009	0.926	1.181	0.795	0.382
15	0.0	0.400	0.600	0.0	0.782	1.087	1.050	0.843	0.667	1.492	0.658	0.070
16	0.0	0.300	0.700	0.0	0.991	1.457	0.781	1.010	0.911	1.400	1.050	-0.179

Table 3. Data Mining Testing Result K-Nearest Neighbor

Based on data mining testing result using K-Nearest Neighbor indicates that classification accuracy of a company that commits fraud is by 87.5% (8 companies which are indicated to commit fraud are accurately predicted by 6 companies). Meanwhile, the classification accuracy of the company with no fraud indication is 100% which means all of which are accurately predicted for no fraud indication. Data mining testing result using K-Nearest Neighbor in performing classification is better than logistic regression. The results of this study are consistent with the findings and Terzi and Şen (2012) shows that the model of Artificial Neural Network (ANN) is more effective than the decision tree model. This study is also consistent with a model of creative accounting. Financial characteristics are smaller in terms of total assets, liquidity and leverage of similar companies.

The study's findings also support Kirkos et al., (2007), Sharma and Panigrahi (2012), Chintalapati et al. (2013) state that the data mining techniques have classification and prediction capabilities that can facilitate the decision-making process in financial matters. Similar findings obtained from research Sharma and Panigrahi (2012) shows that data mining techniques such as logistics model, neural networks (neural networks), the network of trust Bayesian (Bayesian belief network), and a decision tree (decision trees) has been applied most widely for providing the main solution to the problems inherent in the detection and classification of data fraud.

#### 5. Conclusion

Data mining models of both logistic regression and K-Nearest Neighbor have the capability to detect the availability and unavailability of financial fraud indication. The testing result indicates that data mining model of K-Nearest Neighbor has better predictive power than logistic regression model. Detecting using K-Nearest Neighbor is able to detect which company that is accurately classified and not accurately classified whether or not the company shows fraud indication accurately.

## References

Association of Certified Fraud Examiners (ACFE), (2014), Report to The Nation, Austin, Texas.

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

- Amara, I., Amar, A. B., & Jarboui, A. (2013). Detection of Fraud in Financial Statements: French Companies as a Case Study. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 3(3), 40-51.
- Ata, A. & Seyrek, I. (2009) The Use of Data Mining Techniques in Detecting Fraudulent Financial Statements: An Application on Manufacturing Firms, *The Journal of Faculty of Economics and Administrative Sciences*, 14(2) 157-170.
- Asare, S. K., Wright, A., & Zimbelman, M. F. (2015). Challenges Facing Auditors in Detecting Financial Statement Fraud: Insights from Fraud Investigations. *Journal of Forensic & Investigative Accounting*, 7(2).
- Beneish, M. D. (1999). The detection of earnings manipulation. Financial Analysts Journal, 24-36.
- Beneish, M. D., Lee, C., & Nichols, D. C. (2012). Fraud detection and expected returns.
- Beaver, W. H., McNichols, M. F., & Rhie, J. W. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies*, *10*(1), 93-122.
- Beaver, W. H., McNichols, M. F., & Rhie, J. W. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies*, *10*(1), 93-122.
- Chongsirithitisak, T. (2015). Detection of Creative Accounting in Financial Statements by Model The Case Study of Companies Listed on The Stock Exchange of Thailand. *WEI International Academic Conference Proceedings Barcelona, Spain.*
- Crumbley, D.L., Heitger, L.E., Smith, G.S. (2011), Forensic and Investigative Accounting, (5<sup>th</sup> Ed.), 4025 W. Peterson Ave, Chicago.
- Gaganis, C. (2009). Classification techniques for the identification of falsified financial statements: a comparative analysis. *Intelligent Systems in Accounting, Finance and Management, 16*(3), 207-229.
- Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert systems with applications*, 32(4), 995-1003.
- Koroy, T. R. (2009). Pendeteksian Kecurangan (Fraud) Laporan Keuangan oleh Auditor Eksternal [Detection of Fraud Financial Report by the External Auditor]. *Jurnal Akuntansi dan Keuangan*, *10*(1), 22-33
- Kotsiantis, S., Koumanakos, E., Tzelepis, D., & Tampakas, V. (2006). Forecasting fraudulent financial statements using data mining. *International Journal of Computational Intelligence*, *3*(2), 104-110.
- Lenard, M. J., & Alam, P. (2009). An historical perspective on fraud detection: From bankruptcy models to most effective indicators of fraud in recent incidents. *Journal of Forensic & Investigative Accounting*, 1(1), 1-27.
- Ohlson, J.A., 1980, Financial ratios and the prediction of bankruptcy. Journal of Accounting Research 18 (1): 109-131.
- Omoye, A. S. & Eragbhe, E. (2014). Accounting Ratios and False Financial Statements Detection: Evidence from Nigerian Quoted Companies.International Journal of Business and Social Science Vol. 5, No. 7(1).
- Persons, O. S. (1995). Using financial statement data to identify factors associated with fraudulent financial reporting. *Journal of Applied Business Research*, 11(3), 38.
- Prasetyo, E. (2012a). Data Mining Konsep dan Aplikasi Menggunakan Matlab. Andi. Yogyakarta.
- Prasetyo, E. (2012b). Fuzzy K-Nearest Neighbor in every Class untuk Klasifikasi Data Seminar Nasional Teknik Informatika (SANTIKA 2012). pp 57-60.
- Prasetyo, E, (2014), DATA MINING Mengolah Data Menjadi Informasi Menggunakan Matlab, ANDI OFFSET, Yogyakarta.
- Terzi, S., & Şen, İ. K. (2012). Finansal Tablo Hilelerinin Veri Madenciliği Yardimiyla Tespit Edilmesi: Üretim Sektöründe Bir Araştirma. [Determination Of Financial Statements With Data Mining: A Research In The Production Sector]. Muhasebe ve Vergi Uygulamalari Dergisi (MUVU)/Journal of Accounting & Taxation Studies (JATS), 5(2).
- Sharma, A., & Panigrahi, P. K. (2013). A review of financial accounting fraud detection based on data mining techniques. *arXiv preprint arXiv:1309.3944*.
- Suyanto. (2009). Fraudulent Financial Statement Evidence from Statement on Auditing Standard No. 99. *Gadjah Mada International Journal of Business*, 11(1), 117–144.
- Tarjo and Herawati, N. (2015), Application of Beneish M-Score Models and Data Mining to Detect, Financial Fraud, Procedia Social and Behavioral Sciences 211, page 924 930.
- Tarjo, (2008), Effect of Institutional Ownership Concentration and Free Cash Flow to Leverage, Profit Management, Shareholder Value and Cost of Equity Capital, *Dissertation*, (Graduate Program), University of Airlangga, Surabaya, Unpublished.

- Tarjo, (2010), Relevance concentration of ownership and the Debt Policy Affecting Earnings Management, Shareholder Value and Cost of Equity Capital.
- Ujal, M., Amit, P., Hiral, P., & Rajen, P. (2012). Detection of Fraudulent Financial Statement in India: An Exploratory Study. *GFJMR Vol. 4 January-June*.
- Zaki, M., and Theodoulidis (2013), Analyzing Financial Fraud Cases Using a Linguistics-Based Text Mining Approach, (*PhD Research*) Manchester Business School, University of Manchester.