A Review of Text Mining Methods in Fraudulent Financial Statements Detection

BY

Husam A.S Khalifa[1*], Tarek Omran Abdalla Chatanash[2], Baleid Mohammed Aldoukali[3]
[1]Bani Waleed University
[2]College of Technical Sciences -Bani Waleed
[3]University of Tripoli

Abstract:

Financial fraud is one of the most serious threats to the financial industry. Financial statements are basic documents that duplicate a company's economic situation. The fundamental foundations of a decision-making process for financing stakeholders include financial information users such as the general public, creditors, and so on. Financial fraud has severely harmed the long-term growth of financial markets and businesses. The number of financial reporting fraud instances continues to rise. Each incidence is a heavy blow to partners, banks, and financial experts, and it has a big impact on human growth. One of the most important concerns is detecting financial reporting fraud through the use of an active model. The goal of this research is to uncover frauds using various text-mining algorithms in order to protect the public's investments. This research will be beneficial to auditors and financial regulators.

Keywords: Adaptive Crime, Deep Belief Network, Financial statements, Fraud Detection Model Text Mining.

1. Introduction

Technology is now profoundly ingrained in everyone's lives. Almost every action in contemporary life has developed tremendously with technology, from phone conversations to satellites sent into orbit (Patel et al. 2020). The expanding ability to produce and handle information has had a role in the advancement of technology. According to the National Security Agency of the United States, the Internet handles an average of 1826 petabytes a day (Hariri et al. 2019). With the increasing expansion of data and information transmitted through the Internet, it has become vital to control and facilitate the flow of the same (Gandhi et al. 2020).

Text is a prevalent technique of transmitting info in the modern world. Text mining is a multidisciplinary area that covers linguistic processing, information retrieval, web mining, computational linguistics, data extraction, and data mining. Text mining was used to extract automatically organized information from formless and semi-structured articles. It has a great monetary worth.

It is an advanced innovative approach for extracting and knowledge or non-trivial patterns from large collections of formless documents (Sumathy et al. 2013). Text files, flat files, and PDF files are examples of document file types. These files are collected from a variety of sources, including message boards, newsgroups, emails, online chat, SMS, and websites.

Humans can readily solve the challenges and can recognize and apply linguistic patterns to text. However, spelling, contextual meaning, slang, and variances are difficult for machines to handle (Aggarwal et al. 2011). However, in order to understand unstructured data, we must have the linguistic skills as well as the computer capabilities to process text at rapid rates or in big numbers. Text mining assists the computer in performing unstructured data analytic tasks. Financial organizations place a high value on fraud detection.

Financial fraud is a major commercial concern all over the world. Healthcare fraud, mortgage fraud, business fraud, safeties and commodities fraud, financial organization fraud, and other types of fraud are all examples of financial fraud. When fraud prevention fails, fraud detection steps in. Fraud may be unaware, resulting in a failure of fraud prevention. As a result, fraud detection must be utilized frequently (Mostafa. 2013).

When criminals learn about the use of a fraud detection method in a certain procedure, they frequently alter their location. Many thieves are unaware of this type of protection strategy, which makes fraud detection easier (Netzer et al. 2012).

The use of early detection technologies is quite beneficial, yet it is tough to exchange ideas in fraud detection methods. The public description of the fraud detection technique is inadequate. Because thieves would have easy access to all information, it is critical to create more secure systems (Fuller et al. 2011).

Traditional detection approaches are inefficient in terms of cost, security, time consumption, legality, and computing complexity. Some probable methods of crime include assurance claims, credit card businesses, tax reappearance claims, and cell phones. Because of lax security, these types of crimes cost governments and businesses more money (Othman et al. 2015).

Fraud is often known as an adaptable crime. As a result, superior ways of intelligent data inquiry are required to prevent and identify it. They are providing suitable and successful solutions in numerous sectors of fraudulent crimes. These approaches are classified as data mining, text mining, machine learning, statistics, and knowledge discovery in databases (KDD) as a result of data analysis (Dong et al. 2016)..

Financial institutions and banks generate massive amounts of precise transaction data. Some of the most common uses of data mining include trend research, direct marketing, fraud detection, and risk evaluation of new consumers in banks.

The financial area has the highest demand for forecasting. Commodity price forecasting and stock price forecasting can provide substantial profits (Fu et al. 2016). Neural network approaches are employed in prediction, options and bond trading, portfolio management, and acquisitions and mergers. A comprehensive understanding of the nature, consequences and fraudulent financial reporting activities need a competent explanation of financial statement fraud.

The Association of Certified Fraud Examiners (ACFE) defines financial statement fraud as the deliberate misrepresentation or omission of important facts. When weighed against all of the available facts, that is deceiving; it would induce the reader to reconsider (Rawte & G. 2015). Financial statement fraud occurs when a company's executives provide fraudulent financial information. Audits of financial statements ensure that a company's financial reports are free of substantial misrepresentation and fraud. Financial statement fraud is commonly performed in this scheme (Dilla & Robyn. 2015).

In a current hard economy, firms want to be prepared to combat fraudulent activity. A company administration defines reporting fraud as the alteration of financial statement data in order to achieve a false effect (Kanapickienė & Živilė. 2015). Financial statement fraud is a planned and criminal conduct committed by the top management of publicly listed company. There are three types of frauds: rationalization or attitude, opportunity, and pressure or motive.

Every year, financial statement fraud damages the global economy millions of dollars due to the failure of several businesses (West & Maumita. 2016). It is a sort of management fraud that involves the manipulation of financial data. One of the key problems in many businesses is the prevention of financial statement fraud (Kim et al. 2016). Financial reporting fraud includes behaviors such as falsely reporting sales. Reporting money into the current year has largely fit in the prior year. Reporting a cost in the current year or capitalizing expenses in the following year that should be reported (Olszewski. 2016).

Fraudulent observations are commonly used in the detection and identification of fraud. The observations are the consequence of human mistake or attempts to evade fraud prevention mechanisms. Fraud prevention is a method of preventing fraud from occurring in the first place (Albrecht et al. 2015). Feature selection and feature creation activities are carried out in an additional structured spreadsheet layout to describe the amorphous text documents based on these recovered and pre-processed documents. To identify the essential features, feature selection processes need a detailed study of all subdivisions of features (West et al. 2014).

Text Mining techniques are merged after an appropriate collection of characteristics for the applications of the knowledge discovery process, Information Extraction, and Information Retrieval. Classification is commonly used to categorize each item in a batch of data into one of a specified set of classes (Lin et al. 2015). The classification method contracts with software that can organize data substances into groupings. In KDD, this technique is often referred to as the process of identifying motivating, inferred, and inventive data.

Financial statement fraud is more severe in underdeveloped nations, and it is a big economic and social concern (McMahon et al. 2016). A corporation requires a stock exchange to release its financial statements such as revenue, statements of recollected profits, profit and loss account, and cash flow statements on an annual and quarterly basis. A company's true financial health may be replicated by analyzing its financial report. Shareholders in the firm may make informed decisions on funding (Wong & Sitalakshmi. 2015). Money laundering and credit card fraud successes have gradually increased in recent years. It results in the loss of assets owned by individuals and/or businesses. Because of the profit from deception, they endanger the safety of a country that may turn to terrorism. Detecting and combating financial fraud is vital and required. Because of the complexity of trading systems and transactions, detecting fraud is a challenging task. Many fraud detection techniques rely on attribute-value data points derived from transaction data (Throckmorton et al. 2015). The majority of fraud detection research use data mining processes that are focused on structured information and use numerical methodologies.

2. A REVIEW OF FRAUD DETECTION IN FINANCIAL STATEMENTS

SA is one of the most prominent approaches in the area. It has a wide range of uses. Because this approach extracts the underlying views inside textual data, it is also known as

opinion mining (Akaichi et al. 2013). It is widely used in a variety of fields, including e-commerce platforms, blogs, online social media, and microblogs. The motivations for sentiment analysis may be separated into two categories: emotion recognition and polarity detection. The extraction of a collection of emotion labels is the emphasis of emotion detection, whereas polarity detection is more of a classifier-oriented technique with discrete outputs (e.g., positive and negative) (Cambria 2016).

Huang et al (2018) discussed financial fraud detection using inconsistency feature detection. This article describes the CoDetect invention for quickly performing fraud detection on highlight and chart-based similarity matrix. The invention demonstrates yet another approach for uncovering the concept of commercial exercises, ranging from fraud instances to doubting assets. Furthermore, the structure provides an increasingly interpretable technique to detect fraud on a sparse matrix. Exploratory results are tested on both simulated and real datasets, demonstrating that the proposed structure can successfully identify guarded features and fraud embellishments.

Yao et al (2018) demonstrated hybrid data mining plan-based financial statement fraud detection. The fundamental goal of this project is to improve financial fraud detection by combining feature selection with machine learning classification. Re-enactment has clearly proved that irregular forest provides superior execution idea than other frameworks. XGBOOST outperforms the other two feature selection algorithms. Based on the research, two or five components are gradually appropriate in this examination.

Yee et al (2018) presented credit card fraud detection using machine learning in the data mining process. Data mining algorithms were used to assess the patterns and aspects of suspicious and non-suspicious trades using anomalies and standardized data. Aside from that, machine learning techniques were used to predict non-suspicious and suspicious transactions using classifiers. The combination of data mining and machine learning technologies might distinguish genuine and non-genuine interactions by learning data patterns.

Predefined data types are extracted from a written source using information extraction (IE). IE systems primarily seek to identify objects by extracting important information from fragments and then assembling all of the recovered parts into a framework. DiscoTEX (Discovery from Text EXtraction) is one of the fundamental strategies used after extraction to turn structured data into meaningful data in order to find information from it (Salloum et al. 2018).
Named-entity recognition (NER) is used in fiancé to extract specific sorts of data from a page. Customers' transaction order papers may arrive through fax in banking, resulting in extremely varied documents due to the lack of a field template and necessitating effective feature extraction to get a structured document (Emekligil et al. 2016)

Prakash et al (2018) introduced an ATM card fraud detection system based on machine learning techniques. This strategy reduces the number of false alarms and can exhibit accuracy

in detecting fraudulent transactions. In terms of applicability, another method is machine learning. When this approach is integrated into a bank's ATM card fraud detection system, the likelihood of fraudulent transactions may be predicted. Anti-fraud techniques can be implemented to protect banks from massive losses and to reduce risks. The inquiry was motivated by a few concerns in contrast to the fact that we had an alternate misclassification cost.

Jan (2018) established an excellent financial statement fraud detection methodology for controllable progress of financial souks confirmation collected from Taiwan. This model puts the different data mining methodologies into action. In the primary step, an artificial neural network (ANN) and a support vector machine (SVM) approach are designed to screen out crucial elements. In the second step, four types of decision trees are generated for classification purposes, such as chi-square automatic interaction detector (CHAID), classification and regression tree (CART), rapid unbiased efficient statistical tree (QUEST), and C5.0. To create an extremely precise model, non-financial and financial elements are used to detect fake financial information.

Othman (2021) state in her study that the Anti-fraud technology must be backed up by adequate skills, governance, and monitoring. Relying entirely on one-tool technology would leave it incapable of dealing with all types of fraud. The Deloitte Forensic Centre (2008) research, on the other hand, indicated that, despite tremendous effort and time spent on identifying fraud, the rate and quantity of fraud detections have been considerably reduced. According to Zhou and Kapoor (2011), using simple data mining approaches to identify financial statement fraud has a variety of downsides and limitations.

There is a dilemma in that the more executives involved in financial crime are aware of the tools and techniques available for fraud detection, the more likely they are to change their fraud strategies and elude detection, particularly by currently available approaches (Zhou & Kapoor, 2011). New inventive strategies that are both efficient and successful in keeping up with these adaptive or even freshly developing financial frauds are desperately needed.

Future research may look at detecting approaches that might steer the program in response to the unique conditions of a company. When utilizing only past financial statement data to detect fraud, a model may not offer the best prediction (Sharma and Panigrahi, 2012). Since a result, research may consider integrating an investigation of governance aspects, as it has been stated that flaws in corporate governance procedures have resulted in a wave of corporate financial scandals (Fich & Shivdasani, 2007). Exogenous characteristics, which include internal firm-specific elements as well as external ones relating to the economy, industry, and institutional environment, would also enable more accurate financial fraud prediction and detection.

Macro et al (2018) developed financial fraud detection by considering human behavior. The fraud triangle theory is supported by this article, which proposes detecting fraud, a

hypothetical structure that allows to arrange and summarize a group of persons inside a financial institution who submit extortion. Misrepresentation detects method in the strategies for a continuous audit that will be reliable of social event data of professionals presented in user apparatus. According to semantic approaches, the data is linked by the clients behind the inquiry through the accumulation of expressions before being transmitted to a storehouse for subsequent examination. This investigation adds to the realm of cyber security by reducing cases of financial fraud.

Kuldeep et al (2018) predicted credit card fraud detection using lion's share voting and AdaBoost algorithms. In this article, the machine learning technology is used to detect credit card fraud. At first, the usual models are used. Following that, the hybrid techniques are linked, which include the majority voting and AdaBoost processes. To assess model adequacy, a freely available credit card data set is used. A real credit card data set obtained from the banking institution is investigated. Furthermore, noise is added to the data samples and is used to assess the toughness of the methods. The replication findings show that the dominating part voting framework distinguishes fraud situations with greater accuracy than others in the credit card market.

Hajek and Roberto (2017) showed a mining corporation yearly report for the identification of financial wrongdoing in the relative study of machine learning structures. This evaluation acknowledges the improved fraud detection from financial data and business annual reports by generating the consolidated explicit attributes. This task must be completed with a low frequency of dangerous terms in yearly files that may reveal non-fraudulent firms. It is critical to use isolated data to identify fraudulent organizations from widely accessible financial announcements and estimates of sales and profitability.

Stamatis et al (2017) investigated anticipating bogus financial explanations using active learning techniques. This study compares the effectiveness of the AL hypothesis in detecting FFS issues to that of controlled learning strategies. In the related subject, the analog update of antifraud frameworks should be used to enable machine learning instruments. Because of their lack of authority, the objective is to collect reliable datasets that allocate the corresponding proportions of a corporation. It has gone with fraud activities; ways that take use of the presence of a few marked events in order to locate valuable designs from a pool of unlabeled data might be checked incredibly capable.

Chen et al [30] demonstrated a large data-based deception detection plot for financial explanations of business meetings. This inquiry delves at the many aspects of money and economics. Furthermore, estimate big data to improve fraud detection conspire to exactly anticipate financial statement deception for commercial gatherings' financial reports. Furthermore, enhancing investment decision-making and reducing speculative risks and losses benefit creditors and investors.

Dong et al (2017) shown the major influence of social media for financial fraud disclosure in a text mining-based analysis. This evaluation is used in Internet-based life to harness a large amount of user-generated content for fraud detection. To identify amorphous internet-based living substances onto word weights, subjects, and emotion traits, a logical text structure is constructed. The majority of the web-based social networking highlights, moving on to the fraud, which includes the bogus and matched identical businesses. The average accuracy for 10-fold cross-validation is 81.43%. The plan is a major aftereffect of social media material for financial fraud disclosure.

Gupta and Nasib (2012) demonstrated financial fraud detection using text mining. Fraud detection displays a set of financial announcements for both non-fraud and fraud organizations. In financial announcements, these two relationships are followed by pre-processing, which includes a lexical assessment of current material. Since then, a bag of words plot has been used to extract data concealed in the text, resulting in vector spaces for both non-fraudulent and fraudulent relationships.

Kumar and Ravi (2016) conducted a review of text mining applications in the financial area. Their article provides a cutting-edge overview of text mining applications. The text concepts are divided into three categories: client relationship management (CRM), stock market, and FX rate forecast. Some concerns are emerging that are prevalent in optional and consumer advance features. This inquiry might be immensely beneficial to scholars in this sector; the same amount of outstanding issues are highlighted.

Zaki and Babis (2013) suggested an investigating financial fraud cases-based text mining framework based on phonetics. This concept strengthens market observing surveillance frameworks in various financial settings to allow reprocessing of these learning assets, grasp and communicate financial fraud logic activities, and update understanding of financial fraud procedures. A semantic-based text mining approach is constructed from the SEC procedure to acquire data and understand the many market manipulation assortments. It aids in the assessment of fraud during the examination time.

Glancy and Yadav (2011) demonstrated an effective approach for detecting financial statement fraud. CFDM (computational fraud detection model) employs a quantitative approach on textual data. It contains a few strategies for fraud detection that use all of the data present in textual data in general. Existing work reveals setup for perceiving dishonesty in low and high computer-mediated communication (CMC). CFDM provides a systematic technique that is applicable to robotics. Impersonation irritated from 10-K filings based on administration negotiations and inquiry It might distinguish deceptive files from non-fraudulent schemes. When misleading is distrusted, CFDM acts as a selection device. Reproduction was attempted on the MD&A (Management's Discussion and Analysis) from 10-K files, and it was shown to be capable of distinguishing fraudulent submissions from

non-fraudulent schemes. When deception is suspected, CFDM can be useful as a screening method.

3. Problem Statement

On the ground, many data mining methodologies have been used to give decision assistance to stakeholders. Financial statement fraud is a difficult subject for the government and community regulators. Public firms falsify financial statistics not just to cause severe financial losses to all investors, but also to demotivate the financial market. The typical auditing rehearses that may mostly focus on factual examination during the general auditing. Most financial fraud detection efforts in financial statements focus solely on numerical data. Because of purposeful concealment and accounting shenanigans, counterfeit financial data could hardly be distinguished from real data. The usual lack of field for financial fraud statement can be difficult to detect due to the rarity with which it occurs and the fact that it is often committed by skilled persons inside the business who are prepared to sell their deception.

4. Methodology: A Text Mining Approach

Most businesses publish financial statement records that show all of their debts, earnings, costs, and revenue. It may also include some remarks from management depending on business performance. Predictable issues may develop in the future. The various financial statements give a graphical overview of each firm's situation as acquired from the company. The successful company's stock prices are represented by numerous financial documents, from which they determine whether it is eligible for loans or not.

In this research, we propose detecting financial statement fraud using effective feature selection and classification techniques. This approach is divided into four stages: Pre-processing, sampling, feature selection, clustering, and classification are the four steps.

Data correlation analysis and data cleaning, which cleans the noisy data, are conducted during the pre-processing phase. This process also includes data transformation, integration, and reduction. The sample technique is described below, and it involves evaluating a dataset with varied ratios for verification using a hybrid sampling approach. Feature selection is a powerful tool for tackling difficulties in data mining and machine learning technologies. There are three types of feature selection techniques available: filter method, wrapper method, and embedding method. The wrapper system evaluates the chosen characteristics using an analytical presentation of a predetermined learning approach. The filtering system used to evaluate the relevance of attributes is reliant on the particular qualities of the data, and it is independent of any learning methods. An embedded method is used with model learning to embed feature selection, and it is a mutual strategy between the wrapper and filter methods. A feature selection technique has been completed utilizing a filter-based learning scheme. A filter is a means of removing low-scoring characteristics based on the score provided by the

assessment condition. The likelihood of over-fitting is relatively low, which is advantageous for large-scale data. Assess the significance of attributes using trustworthy data features that are independent of learning techniques.
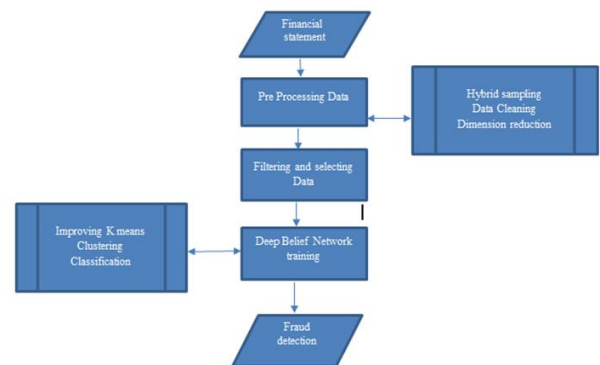

Figure.1: Financial fraud detection Approach

After a feature selection procedure, the clustering process uses enhanced K-means, and the output is utilized as the classification process's training and validation set (Fig.1). Within the deep belief network, the optimization procedure minimizes the goal function (Error Rate). As a result, improved deep learning can improve the accuracy of fraud detection.

Financial statement fraud detection may be evaluated using datasets such as COMPUSTAT, PagSeguro, the Athens stock exchange, and 1610 firm-year samples from publicly traded businesses in the United States, as well as any new datasets that become accessible in the near future. These datasets may be used for 70% training and 30% testing.

The suggested method's performance may be evaluated using statistical metrics such as recall, sensitivity, F-measure, specificity, and classification accuracy. The results of the simulation will be examined and compared to the previous categorization approach.

5. Conclusion

To improve accuracy, the categorization procedure employs an improved deep belief network (DBN). Furthermore, it leads to a cheaper computational cost. A DBN typically has three layers of RBMs (Restricted Boltzmann Machines), with the first layer serving as the visible layer and the second serving as the concealed layer. The last RBM's hidden layer uses the output of the DBN as the visible layer, which is treated as a single unit. The following approaches are used in the categorization procedures listed above:

1. To examine several cutting-edge strategies for detecting financial fraud.
2. To resolve the disputes between the various financial statement frauds.
3. Examine several unsupervised learning techniques for financial statement grouping.
4. Improve financial statement fraud classification accuracy by extracting many optimum features through the feature selection procedure.

The suggested method's performance may be evaluated using statistical metrics such as recall, sensitivity, F-measure, specificity, and classification accuracy. The simulation results may be analyzed and compared to previous categorization algorithms.

References

1. Akaichi J, Dhouioui Z, López-Huertas Pérez MJ (2013) Text mining Facebook status updates for sentiment classification. In: 2013 17th international conference on system theory, control, and computing (ICSTCC), Sinaia, 2013, pp 640–645. https://doi.org/10.1109/ICSTCC.2013.6689032

2. Aggarwal, Charu C., and Haixun Wang. "Text mining in social networks." In Social network data analytics, pp. 353-378. Springer, Boston, MA, 2011.

3. Albrecht, Chad, Daniel Holland, Ricardo Malagueño, Simon Dolan, and Shay Tzafrir. "The role of power in financial statement fraud schemes." Journal of Business Ethics 131, no. 4 (2015): 803 813.

4. Cambria E (2016) Affctive computing and sentiment analysis. IEEE Intell Syst 31(2):102–107. https://doi.org/10.1109/ MIS.2016.31

5. Chen, Yuh-Jen, and Chun-Han Wu. "On Big Data-Based Fraud Detection Method for Financial Statements of Business Groups." In Advanced Applied Informatics (IIAI-AAI), 2017 6th IIAI International Congress on, pp. 986-987. IEEE, 2017.

6. Dilla, William N., and Robyn L. Raschke. "Data visualization for fraud detection: Practice implications and a call for future research." International Journal of Accounting Information Systems 16 (2015): 1-22.

7. Dong, Wei, Shaoyi Liao, and Liang Liang. "Financial Statement Fraud Detection using Text Mining: A Systemic Functional Linguistics Theory Perspective." In PACIS, p. 188. 2016.

8. Dong, Wei, Shaoyi Liao, Yujing Xu, and Xiaoqian Feng. "Leading effect of social media for financial fraud disclosure: A text mining based analytics." (2016).

9. Emekligil E, Arslan S, Agin O (2016) A bank information extraction system based on named entity recognition with CRFs from noisy customer order texts in Turkish. In: Knowledge engineering and semantic web, pp 93–102

10. Fich, E. M., and Shivdasani, A. (2007). Financial fraud, director reputation, and shareholder wealth. Journal of Financial Economics, 86(2), 306-336.

11. Fu, Kang, Dawei Cheng, Yi Tu, and Liqing Zhang. "Credit card fraud detection using convolutional neural networks." In International Conference on Neural Information Processing, pp. 483-490. Springer, Cham, 2016.

12. Fuller, Christie M., David P. Biros, and Dursun Delen. "An investigation of data and text mining methods for real-world deception detection." Expert Systems with Applications 38, no. 7 (2011): 8392-8398.

13. Gandhi M, Kamdar J, Shah M (2020) Preprocessing of non-symmetrical images for edge detection. Augment Hum Res 5:10. https://doi.org/10.1007/s41133-019-0030-5

14. Glancy, Fletcher H., and Surya B. Yadav. "A computational model for financial reporting fraud detection." Decision Support Systems 50, no. 3 (2011): 595-601.

15. Gupta A, Bhatia P, Dave K, Jain P (2019) Stock market prediction using data mining techniques. In: 2nd international conference on advances in science and technology, pp 1–5

16. Gupta, Rajan, and Nasib Singh Gill. "Financial statement fraud detection using text mining." Editorial Preface 3, no. 12 (2012): 189-191.

17. Hajek, Petr, and Roberto Henriques. "Mining corporate annual reports for intelligent detection of financial statement fraud–a comparative study of machine learning methods." Knowledge-Based Systems 128 (2017): 139-152.

18. Hariri RH, Fredericks EM, Bowers KM (2019) Uncertainty in big data analytics: survey, opportunities, and challenges. J Big Data. https://doi.org/10.1186/s40537-019-0206-3.

19. Huang, Dongxu, Dejun Mu, Libin Yang, and Xiaoyan Cai. "CoDetect: Financial Fraud Detection with Anomaly Feature Detection." IEEE Access 6 (2018): 19161 -19174.

20. Jan, Chyan-long. "An Effective Financial Statements Fraud Detection Model for the Sustainable Development of Financial Markets: Evidence from Taiwan." Sustainability 10, no. 2 (2018): 513.

21. Kanapickienė, Rasa, and Živilė Grundienė. "The model of fraud detection in financial statements by means of financial ratios." Procedia-Social and Behavioral Sciences 213 (2015): 321 -327.

22. Karlos, Stamatis, Georgios Kostopoulos, Sotiris Kotsiantis, and Vassilis Tampakas. "Using Active Learning Methods for Predicting Fraudulent Financial Statements." In International Conference on Engineering Applications of Neural Networks, pp. 351 -362. Springer, Cham, 2017.

23. Kim, Yeonkook J., Bok Baik, and Sungzoon Cho. "Detecting financial misstatements with fraud intention using multi-class cost-sensitive learning." Expert systems with applications 62 (2016): 32-43.

24. Kumar, B. Shravan, and Vadlamani Ravi. "A survey of the applications of text mining in financial domain." Knowledge-Based Systems 114 (2016): 128-147.

25. Lin, Chi-Chen, An-An Chiu, Shaio Yan Huang, and David C. Yen. "Detecting the financial statement

fraud: The analysis of the differences between data mining techniques and experts' judgments." Knowledge-Based Systems 89 (2015): 459-470.

26. McMahon, Richard, Diana Pence, Linda Bressler, and Martin S. Bressler. "New Tactics in Fighting Financial Crimes: Moving Beyond the Fraud Triangle." Journal of Legal, Ethical & Regulatory Issues 19, no. 1 (2016).

27. Mostafa, Mohamed M. "More than words: Social networks' text mining for consumer brand sentiments." Expert Systems with Applications 40, no. 10 (2013): 4241 -4251.

28. Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko. "Mine your own business: Market-structure surveillance through text mining." Marketing Science 31, no. 3 (2012): 521-543.

29. Olszewski, Dominik. "Fraud detection using self-organizing map visualizing the user profiles." Knowledge-Based Systems 70 (2014): 324-334.

30. Othman, Intan Waheedah. " Financial Statement Fraud: Challenges and Technology Deployment in Fraud Detection " International Journal of Accounting and Financial Reporting, Vol. 11, No. 4 (2021).

31. Othman, Rohana, Nooraslinda Abdul Aris, Ainun Mardziyah, Norhasliza Zainan, and Noralina Md Amin. "Fraud detection and prevention methods in the Malaysian public sector: Accountants' and internal auditors' perceptions." Procedia Economics and Finance 28 (2015): 59-67.

32. Patel D, Shah Y, Thakkar N, Shah K, Shah M (2020a) Implementation of artificial intelligence techniques for cancer detection. Augment Hum Res. https://doi.org/10.1007/s41133-019-0024-3\

33. Prakash, B., G. Venu Madhava Murthy, P. Ashok, B. Pavan Prithvi, and S. Sai Harsha Kira. "ATM Card Fraud Detection System Using Machine Learning Techniques." (2018).

34. Randhawa, Kuldeep, Chu Kiong Loo, Manjeevan Seera, CHEE PENG Lim, and Asoke K. Nandi. "Credit card fraud detection using AdaBoost and majority voting." IEEE ACCESS 6 (2018): 14277-14284.

35. Rawte, Vipula, and G. Anuradha. "Fraud detection in health insurance using data mining techniques." In Communication, Information & Computing Technology (ICCICT), 2015 International Conference on, pp. 1-5. IEEE, 2015.

36. Salloum S, Mostafa A, Monem A, Shaalan K (2018) Using text mining techniques for extracting information from research articles. https://doi.org/10.1007/978-3-319-67056-0_18

37. Sánchez, Marco, Jenny Torres, Patricio Zambrano, and Pamela Flores. "FraudFind: Financial fraud detection by analyzing human behavior." In Computing and Communication Workshop and Conference (CCWC), 2018 IEEE 8th Annual, pp. 281 -286. IEEE, 2018.

38. Sharma, A., & Panigrahi, P. K. (2012). A review of financial accounting fraud detection based on data mining techniques. International Journal of Computer Applications, 39(1), 37-47.

39. Sumathy, K. L., and M. Chidambaram. "Text mining: concepts, applications, tools, and issues an overview." International Journal of Computer Applications 80, no. 4 (2013).

40. Throckmorton, Chandra S., William J. Mayew, Mohan Venkatachalam, and Leslie M. Collins. "Financial fraud detection using vocal, linguistic and financial cues." Decision Support Systems 74 (2015): 78-87.

41. Yao, Jianrong, Jie Zhang, and Lu Wang. "A financial statement fraud detection model based on hybrid data mining methods." In 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), pp. 57-61. IEEE, 2018.

42. Yee, Ong Shu, Saravanan Sagadevan, and Nurul Hashimah Ahamed Hassain Malim. "Credit Card Fraud Detection Using Machine Learning as Data Mining Technique." Journal of Telecommunication, Electronic and Computer Engineering (JTEC) 10, no. 1 -4 (2018): 23-27.

43. West, Jarrod, and Maumita Bhattacharya. "Some Experimental Issues in Financial Fraud Mining." In ICCS, pp. 1734-1744. 2016.

44. West, Jarrod, Maumita Bhattacharya, and Rafiqul Islam. "Intelligent financial fraud detection practices: an investigation." In International Conference on Security and Privacy in Communication Systems, pp. 186-203. Springer, Cham, 2014.

45. Wong, Shirley, and Sitalakshmi Venkatraman. "Financial accounting fraud detection using business intelligence." Asian Economic and Financial Review 5, no. 11 (2015): 1187.

46. Zaki, Mohamed, and Babis Theodoulidis. "Analyzing financial fraud cases using a linguistics-based text mining approach." (2013).

47. Zhou, W., & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. Decision Support Systems, 50(3), 570-575. https://doi.org/10.1016/j.dss.2010.08.007