Financial Statement Audit using Support Vector Machines, Artificial Neural Networks and K-Nearest Neighbor:
An Empirical Study of UK and Ireland

Aram Khalaf Nawaiseh, Maysam F. Abbod, Takebumi Itagaki

Department of Electronic and Computer Engineering
College of Engineering, Design and Physical Sciences, Brunel University London, London, UK.

aram.nawaiseh@brunel.ac.uk; maysam.abbod@brunel.ac.uk; take.itagaki@brunel.ac.uk

Abstract **- Initial applications of big data analytics such as data mining techniques in auditing remain pioneering, and more research is needed on attributes to develop predicted models using data mining analytics on financial statement auditing. This study explores data mining abilities based on Support Vector Machines (SVM), Artificial Neural Networks (ANN), and K-Nearest Neighbor (KNN) as predictive classification models for financial statement audit. Empirical results showed that ANN and SVM techniques achieved higher average accuracy rate and outperformed KNN in correctly classifying healthy companies. However, ANN had the lowest rate of Type I error, indicating better ability in classifying healthy companies' compeer to other techniques. SVM had better performance in terms of fewer incorrect classification of qualifed companies into unqualified class, with the lowest rate of Type II error. This study demonstrates the superiority of ANN and SVM in predictive classification of correct auditing opinions. This is a pioneering auditing study using big data financial information to investigate attributes of developing prediction models.**

Keywords - financial statement audit, big data, data mining, computer assisted audit tools, support vector machine, artificial neural network, k-nearest neighbour

I. INTRODUCTION

Financial statement audit involves searching for evidence relevant to management assertions of firms' financial and non-financial data, then evaluating and attesting such evidence objectively to determine the credibility of such assertions in relation to accounting criteria. This process is doing by independent auditors, using their experience and skills to report on the quality of financial statements and reliability assurance, free of fraud or error [1]. Rapid technological progress has proliferated fraudulent business schemes [2]. Financial statement fraud has adverse impacts on companies and investors, undermining the financial statements' credibility [3]. Technological advances and new software enable auditors to perform audit data analysis and collect evidence through new methods, increasing the complexity of traditional auditing in terms of collecting, tabulating, and storing accounting data [4].

Technological advances and new software also empower auditors to perform audit data analysis, and to collect big data (BD) relevant to the auditing process from different sources, which can manufacture audit evidence applied in the analytical procedures, review testing, risk assessment, and assurance on the reliability financial statement, through new methods [5]. Previous studies like [3], [6] have shown that statistical and machine learning-based technologies are effective at detecting and preventing fraud, but usually fraudsters can adapt and discover other ways of circumvention.

Accountancy and auditing have seen major impacts of data and data analytics on their roles, but initial applications of BD analytics (BDA) such as data mining in the auditing profession are still in their infancy, and auditors are still relative novices at deploying BDA in audit procedures. However, pressures are mounting on the auditing field to improve expressed judgments on financial statements. More research is needed to illustrate and investigate the behavioral implications of using BDA and data mining in financial statement auditing, and more attributes must be identified to develop predictive models using large data sets, with improved accuracy based on the latest BDA research on auditing.

The main aim of this study is to explain the probable ability of data mining as a classification tool in financial statement auditing. It endeavors to evolve an accurate prediction model utilizing variables from non-financial and financial ratios to predict audit opinion by utilizing data mining techniques. This study is a pioneering effort in using BD financial information to develop a prediction model able to correctly predict audit opinions.

II. RELATED WORK

This section gives an overview of previous studies on computer assisted audit tools (CAAT) (Section A), BD and auditing (Section B), and data mining (Section C).

A. Computer Assisted Audit Tools

CAAT comprises auditing technology utilized via both external and internal auditors to aid in the accomplishment of financial audit processes like audit testing, and processes related to extracting and analytics data, implementing application control tests, audit testing, and processing audit data in information systems. CAAT utilizes computer-aided procedures for audit information located on information systems. Its main goal is to facilitate the process of financial auditing [7]. CAAT tools can be categorized into four broad classes: operating system; data analytics software; software/utility and code tool of test; and network security estimate utility/software [8].

CAAT has advantages for auditors in reducing audit cost; enhancing the quality, effectiveness, and efficiency of auditing; and supporting timely reporting [7]. CAAT plays a leading role in guaranteeing audit report fineness. It is utilized to earn and produce audit evidence and data, and it is effective in checking firm transactions. It enables auditors to test the whole population rather than a sample, consequently increasing the accuracy of audit testing and outcomes [9]. CAAT represents an innovative method of data analytics, fundamentally due to authorizing the manipulation of massive data registration with no considerable added cost [8].

B. Big Data and Auditing

BD is a new phenomenon whose existence is caused by the pooling of vast volumes of complicated and heterogeneous data generated anytime, anywhere, of increasing pertinence and value in the epoch of the Internet of Things. However, despite its fundamental importance, BD lacks a clear and universal definition. Many researchers define it based on properties known as the seven V's: variability, variety, volume, velocity, veracity, visualization, and value [10], [11]. BD is essentially a method of data analysis made possible by recent technological advances enabling the capturing of variable and complex (semi-structured, unstructured, and structured data) with high velocity, facilitating its analysis, administration, distribution, and storage [11].

A conceptual sample of the seven V attributes of BDA was evolved to acquire deeper comprehension of the practices and strategies of high-frequency trading in financial markets [10]. Fig. 1 shows the developer model to integrate the BD seven V attributes according to additional explanation of BD among high-frequency trading companies in the financial market, dividing the seven V's to BD to manipulate and store data, accessing fast data in real-time, using big data computing ability to treat senior quantities of information. The model indicates that high-frequency trading companies give priority to these characteristics based on their hold market strategy.

BDA is applied to inspect, treat, convert and model data from numerous different sources and identify and alert users to patterns it identifies, suggesting preliminary conclusions to support decision making [12]. Presently, auditors rely on both structured and unstructured data in forming judgments on the credibility of financial statements. BDA adds more unstructured, non-traditional data to the evidence currently utilized by auditors, and it can disseminate auditing documentation beyond conventional corporate databases to integrate automated inventory and supply chain data (among other types) from smartphone and sensor devices etc. It thus opens up a vast volume of unstructured and unformatted data accessed using modern technology, which makes financial statement auditing more complicated [13], [14].

Conceptual Model of Big Data in High-Frequency Trading
(BG = Big Data; FD = Fast Data; BC = Big Compute

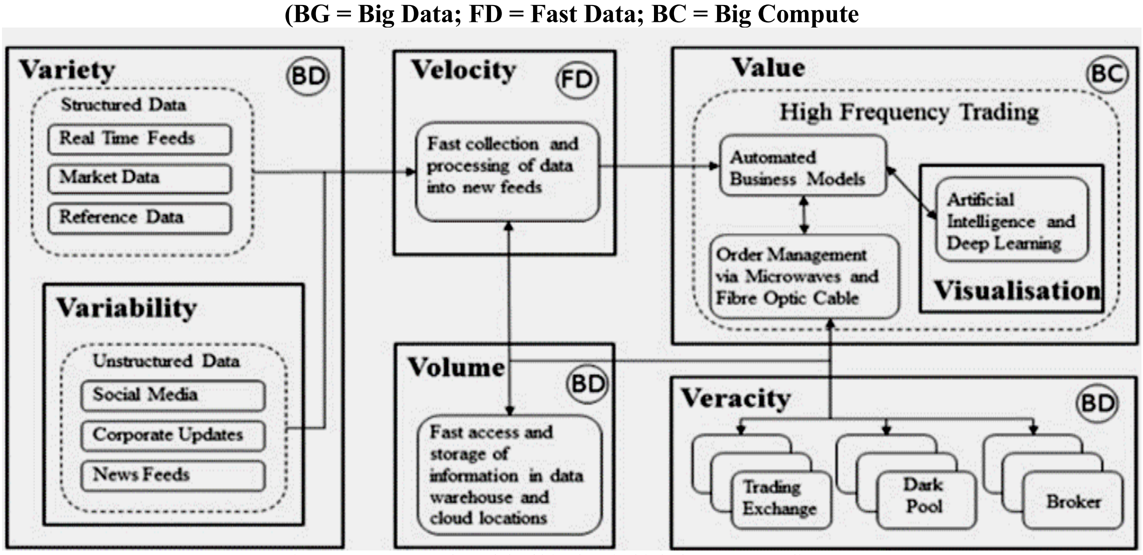

Figure 1. Conceptual model of big data in high-frequency trading [10]

Even though the most recent auditing standards have increased the responsibility of auditors to detect if there is any fraud in financial statements, it remains a problematic task for the accounting and auditing profession due to the rapid enhancements of technology outpacing the capabilities of auditors to utilize available tools effectively, and the advanced knowledge and skills of actual fraudsters [15]. [16]. Non-financial information provided by BD can be used in the process of identifying and detecting fraud [17]. The fraud triangle determines size rationalization, incentives, and opportunities, involving qualities of uncertainty that may not be easily recognized by the figures in financial statements. Consequently, BD analysis is beneficial to relieve the problem of how to detect fraud and improve the validation of financial statements. Through the integration of semi-structured and unstructured data and supplying more trustworthy outcomes that may help internal auditors in determining gaps or new conditions acquired by individual networks, fraudulent activity and errors can be detected more effectively and efficiently [18].

Nevertheless, new auditing information from BD is worthless for auditors without effective analysis [19], and the value of BD itself is defined by the analysis performed therewith [20]. Because of this, different data analytical instruments, like data mining, are considered beneficial to auditors to analyze data.

C. Data Mining

Data mining (DM) has been increasingly utilized in advanced auditing in audit phases, affecting audit opinion outcomes [21]. DM has played a significant role in financial statement fraud detection by extracting and uncovering unobserved realities from huge volumes of data [22]. DM uses a set of mathematical, machine learning, statistical, and artificial intelligence processes to elicit valuable information from data, and to determine interesting manners in databases that may be utilized in the decision-making process. The purpose of the DM process is to secure beneficial and unauthorized information from data stockpiled in enormous data warehouses [6].

DM techniques have been effective in decision-making processes and enhancing decision support tools as financial indicators, through the aid of an evolutionary methodology that enables auditors to discover institutional manipulations from red flags based on experience. This knowledge enhances the decision support framework concerning classifying financial statements as false or true. Auditors can classify the credibility of companies' financial statements using the learning process of DM rating techniques [23].

In recent years, studies have discussed the effect and challenge of using data mining techniques on accounting, especially in auditing, but very few researchers have started analyzing DM data to understand impacts on audit opinion, financial statement fraud detection, and bankruptcy prediction[6], [24], [25]. DM discriminatory power to predict bankruptcy has been explored by combining seven characteristics of financial ratios with five characteristics of corporate governance indicators, using data from 478 companies listed on the Taiwan Stock Exchange for ten years [25]. Five techniques were used to enhance prediction techniques: regression tree, naïve Bayes, SVM, multilayer perceptron methods, and KNN. The outcome illustrated that, compared with techniques based on financial ratio alone, a combination of CGIs with FRs may enhance prediction model performance.

This study uses an extensive sample of multiple data sets from around 500 companies, using 30 different characteristics of financial ratios.

III. RESEARCH METHODOLOGY

This part presents the prediction models used to analyze the data set (Section A), data collection (Section B), and the Criteria of comparing techniques' results (Section C).

A. Prediction Models

This section presents three data mining techniques commonly utilized for classification in audit opinion prediction: Support Vector Machines (SVM), Artificial Neural Networks (ANN), and K-Nearest Neighbor (KNN).

A1. Support Vector Machines

SVM is a classification modelling method based on statistical learning theory related to learning algorithms utilized to analyze data in regression and classification; they are thus simple and sufficient to be examined mathematically [26]. A linear optimal hyperplane splits data, with a maximum margin between the hyperplane and the nearest point, to enable binary classification using mapping input vectors into the high dimensional characteristic distance, through nonlinear transformation. Training points that are near to the extreme margin hyperplane launch a support vector. All other training points are not relevant for setting the dual-class frontiers, after which the SVM model classifies the unknown data set into output classes [27].

Researchers have used SVM due to two main advantages [28]: (a) SVM solution is optimal, unique, and universal, since the SVM training achieved through resolving the linearly constrained quadratic issues; (b) SVM is lean on the principle of the structural risk reduction, which means that this kind of classifier reduces an upper bound of the actual risk, while other classifiers decrease the empirical risk. Because of this, a substantial number of studies have explored its application and theory, and financial applications have increasingly utilized SVM for time series prediction, insurance fraud detection, and credit rating. In light of this, previous researches reported that SVM has successful application in these areas and performance is comparable or even preferable to other traditional classifiers, like logistic regression and discriminate analysis [26].

A2. Artificial Neural Networks

ANN is a learning algorithm mechanism based on the human brain system. It can be applied to model a mathematical non-linear statistical data in massive volumes. As illustrated in Fig. 2, ANN processing has vast parallel dispensers and is composed of simple handling units called neurons. Each communication is connected with a numeral value summon weight [22]. The output unit of ANN picks the weighted totality of the outputs from units in a former stratum. Input is pushed from input nodes through the hideaway stratum to output nodes. This ranking model may easily adjust through changing the weights in former stratums. The ANN multi-layer feed utilizes a backpropagation algorithm to adjust weightings, decreasing mean squared error between the actual output value and the network's prediction value [24].

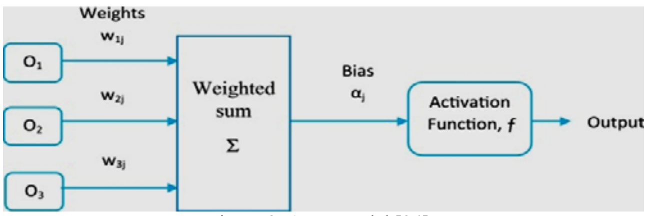

Figure 2. ANN model [24]

ANN is the most attractive substitute tool in many research fields, particularly economics and finance. It has the physical ability to stock and discover underlying functional relevance between inputs and outputs, and implement tasks like classification, control, pattern recognition, modelling, evaluation, and prediction. ANN can be particularly useful when employed to calculate accurate solutions for noisy, complex, irrelevant, or partial data [29]. It uses interactions between numerous different variables that are substantially correlated, frequently supposed to be nonlinear, unclearly related, and complicated to be illustrated using statistical models [30].

A3. K-Nearest Neighbor

KNN technique includes objects with N characteristics, considered to be the N distance space. Every object is one spot in the N distance space. KNN introduces the resemblance metric for the object. When obscure observations are given, KNN classifies the new items by comparing every new item with existing ones through utilizing the distance measurement by the calculated distance between every item in a sampling set, after which it determines the closest K neighbor to the unknown observation. This K case is the eponymous KNN of a new observation. The classifier specifies a new observation to the most popular class between a KNN [31], [32]. Different areas of financial and accounting studies have utilized KNN analytics, like audit decision-making and the detection of financial statement fraud, and it is one of the significant classifier algorithms employed in fraud detection [33], [34].

Researchers in [34] examined the competence of the KNN model to improve methods for auditors' opinions, as opposed to methods sophisticated with logit and discriminate analytics, using a sample of UK firms' financial statements. The researchers compared between discriminate and logit analyses and KNN model and determined that KNN is more effective and efficient in terms of average classification accuracy. They collected two data sets, including financial data (financial ratio, and annual statements such as financial statement, income statement, cash flow statement, and equity statement) and audit opinion from all industries in British and Irish companies using the Financial Analysis Made Easy Database (FAME) software, which contains data for about 11 million firms in the studied countries. This software has up to ten years of companies' data, including financial information (annual report, audit detail, etc.), calculated financial ratio related to annual statement for each company, and descriptive and information about firms [35].

A total of 49 independent variables were chosen that consist of financial and non-financial variables, and the dependent variable is audit opinion: qualified (firms do not have credibility in the financial statement) and unqualified (companies do not have any fraud or misstatements). Two data sets were used, the first for one year, and the second for five years. Table I presents an overview of the sample of companies used in this study.

TABLE I. DATA SETS

	Data Sets		
	No. of Companies	***No. of Qualified Companies***	***No. of Unqualified Companies***
Data set 1	23,211	4,211	19,000
Data set 2	30,199	7,199	23,000

B. Comparing Techniques Criteria

In this study, the performance of the developed models was inspected using Type I and Type II errors and accuracy rate (adopted from previous studies in accounting and finance, which used this two-evaluation measurement to measured model performance). The following equations are used to calculate accuracy (1), Type I error (2) and Type II error (3):

$$\text{Average Accuracy} = TP+TN/\ (TP+FN+TN+FP) \quad (1)$$
$$\text{Type I error} = FP/\ (TN+FP) \quad (2)$$
$$\text{Type II error} = FN/\ (TP+FN) \quad (3)$$

Therefore, *true* and *false* refer to the assigned classifications being correct or incorrect. True negative refers to correctly classified failed companies, while true positive

DOI 10.5013/IJSSST.a.21.02.07 ISSN: 1473-804x online, 1473-8031 print

ratio refers to correctly classified healthy companies. Table II shows the classification test.

TABLE II. CLASSIFICATION TEST

Actual Class (%)	Predicted Classifier (%)	
	Qualified	*Unqualified*
Qualified	**True positive (TP)**	**False negative (FN)**
Unqualified	**False Positive (FP)**	**True Negative (TN)**

IV. EXPERIMENTAL RESULTS

This study applied SVM, KNN, and ANN techniques using MATLAB. The models were safeguarded from overfitting using cross-validation method, partitioning the data set into folds, and estimating accuracy for each fold.

A. Support Vector Machines Model Results

Classification learning application tool tested different types of SVM modelling techniques. Quadratic SVM technique was selected as it produced the highest prediction rates. Tables III and IV show the classification and valuation metrics' results of SVM technique performance for the two data sets.

TABLE III. CLASSIFICATION RESULTS OF SVM MODEL

Classifier	TP	TN	FP	FN
Data set 1	81%	99%	1%	19%
Data set 2	76%	99%	1%	24%

TABLE IV. VALUATION METRICS RESULTS OF SVM MODE

	Average accuracy	Type I error	Type II error
Data set 1	95.9%	1%	19.%
Data set 2	95.4%	1%	24%

B. Artificial Neural Networks Model Results

ANN utilized the neural pattern recognition app (nprtool) to select data and divide it into three data sets (training, testing, and validation), and to create, train, and evaluate network performance utilizing confusion matrices and cross-entropy. 50 hidden layers were defined (Fig. 3), which were divided into three partitions for training (70%), testing (15%), and validation (15%). Tables V and VI show the classification and valuation metrics results of ANN model performance for the two data sets.

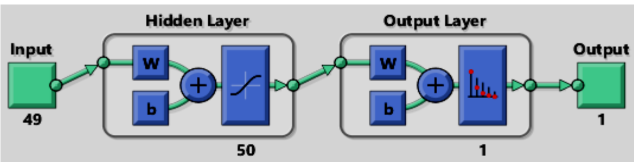

Figure 3. Neural Networks architecture.

TABLE V. CLASSIFICATION RESULTS OF ANN MODEL

Classifier	TP	TN	FP	FN
Data set 1	79.7%	99.5%	0.5%	20.3%
Data set 2	74.1%	99.2%	0.8%	25.9%

TABLE VI. VALUATION METRICS RESULTS OF ANN MODEL

	Average Accuracy	Type I error	Type II error
Data set 1	95.9%	0.5%	20.3%
Data set 2	95.2%	0.8%	25.9%

C. K-Nearest Neighbor Model Results

Cosine KNN was chosen as it had the highest prediction rates, determining 10 neighbors. Tables VII and VIII show the classification and valuation metrics results of KNN model performance for the two data sets.

TABLE VII. CLASSIFICATION RESULT OF KNN MODEL

Classifier	TP	TN	FP	FN
Data set 1	72%	99%	1%	28%
Data set 2	64%	99%	1%	36%

TABLE VIII. VALUATION METRICS RESULTS OF KNN MODEL

	Average Accuracy	Type I error	Type II error
Data set 1	93.8%	1%	28%
Data set 2	93.4%	1%	36%

D. Results Comparison

Looking more closely at the Tables IV, VI, and VIII of valuation metrics results for ANN, SVM, and KNN techniques, it can be seen that the lowest rate of Type I error is achieved by ANN, with a better ability to classify healthy companies than SVM and KNN. SVM had better performance in terms of fewer incorrect classification of qualified companies into unqualified class, with the lowest rate of Type II error. Further comparison of the valuation metrics results reveals that ANN and SVM techniques achieved 95.9% average accuracy to outperform correctly classifying the companies, but the KNN technique had only 93.8%.

Valuation metrics results indicate that ANN and SVM can outperform KNN due to: ANN having the ability to model nonlinear systems, because the data sets have nonminority, with a lot of normalized and data missing, which ANN copes well with; the data set having unbalanced data between the number of qualified and unqualified companies. Consequently, SVM outperforms due to its good features in modelling unbalanced data.

V. CONCLUSION

Financial statement auditing is a process of searching for and evaluating evidence objectively to determine credibility. Opinion audits can be divided into qualified (firm annual statements are not credible) and unqualified (firm annual statements are credible). This study explained the probable ability of data mining based on SVM, ANN, and KNN as classification tools in financial statement audit. It evolved an accurate prediction model utilizing 49 variables from non-financial and financial ratios to predict audit opinion (qualified and unqualified) utilizing data mining techniques. This study is a pioneering effort in using BD financial information to develop a model able to predict the right audit opinion. However, challenges were faced in data collection and cleaning, dealing with noise, and unbalanced and missing data in the data sets.

The empirical results indicate that the ANN and SVM techniques achieve higher average accuracy, outperforming KNN in correctly classifying companies. However, ANN had the lowest rate of Type I error, indicating superior ability in classifying healthy companies. SVM had better performance in terms of fewer incorrect classification of qualified companies into unqualified class, with the lowest rate of Type II error. Future work will address issues utilizing other data mining techniques, like deep learning, in order to improve classification and audit opinions.

REFERENCES

[1] I. Gray and S. Manson, *The Audit Process: Principles, Practice and Cases*, 5th ed. Mason, OH: South-Western, 2011.

[2] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Statist. Sci.*, vol. 17, no. 3, 2002, pp. 235–255.

[3] W. Zhou and G. Kapoor, "Detecting evolutionary financial statement fraud," *Decis. Support Syst.*, vol. 50, no. 3, 2011, pp. 570–575.

[4] S. Buchanan and F. Gibb, "The information audit: Role and scope," *Int. J. Inf. Manage.*, vol. 27, no. 3, 2007, pp. 159–172.

[5] B. A. Sirois and K. S. Savovska, *Audit Data Analytics: Opportunities and Tips*, International Federation of Accounting, 2017. Accessed on: Feb. 5, 2020. [Online]. Available: https://www.ifac.org/global-knowledge-gateway/audit-assurance/discussion/audit-data-analytics-opportunities-and-tips

[6] M. Albashrawi, "Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015," *J. Data Sci.*, vol. 14, no. 3, 2016, pp. 553–569.

[7] E. M. Mansour, "Factors affecting the adoption of computer assisted audit techniques in audit process: Findings from Jordan," *Bus. and Econ. Res.*, vol. 6, no. 1, 2016, pp. 1–24.

[8] I. Pedrosa and C. J. Costa, "Computer assisted audit tools and techniques in real world: CAATT's applications and approaches in context," *Int. J. Comput. Inf. Syst. and Ind. Manage. Appl.*, vol. 4, 2012, pp. 161–168.

[9] J. B. Omonuk and A. A. Oni, "Computer assisted audit techniques and audit quality in developing countries: Evidence from Nigeria," *J. Internet Banking and Commerce*, vol. 20, no. 3, 2015, pp. 1–17.

[10] J. J. Seddon and W. L. Currie, "A model for unpacking big data analytics in high-frequency trading," *J. Bus. Res.*, vol. 70, 2017, pp. 300–307.

[11] X. Wu, X. Zhu, G.- Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, 2014, pp. 97–107.

[12] J. Zakir, T. Seymour, and K. Berg, "Big data analytics," *Issues in Inf. Syst.*, vol. 16, 2015, pp. 81–90.

[13] J. Zhang, X. Yang, and D. Appelbaum, "Toward effective big data analysis in continuous auditing," *Accounting Horizons*, vol. 29, no. 2, 2015, pp. 469–476.

[14] P. Goes, "Big data and IS research," *MIS Quart.*, vol. 38, no. 3, 2014, pp. iii-viii.

[15] S. Wang, "A comprehensive survey of data mining-based accounting-fraud detection research," in *Int. Conf. on Intell. Comp. Tech. and Automat.*, 2010, pp. 50–53, doi: 10.1109/ICICTA.2010.831.

[16] S. E. Kaplan, K. Pany, J. Samuels, and J. Zhang, "An examination of anonymous and nonanonymous fraud reporting channels," *Advances in Accounting*, vol. 28, no. 1, 2012, pp. 88–95.

[17] Y. Tang, Y.- Q. Zhang, N. Chawla, and S. Krasser, "Correspondence SVMs modeling for highly imbalanced classification," *IEEE Trans. Syst., Man, Cybern. Syst., Part B*, vol. 39, no. 1, 2009, pp. 281–288.

[18] D. Appelbaum, "Securing big data provenance for auditors: The big data provenance black box as reliable evidence," *J. Emerg. Tech. in Accounting*, vol. 13, no. 1, 2016, pp. 17–36.

[19] J. D. J. Warren, K. C. Moffitt, and P. Byrnes, "How big data will change accounting," *Accounting Horizons,* vol. 29, no. 2, 2015, pp. 397–407.

[20] M. Alles and G. L. Gray, "Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors," *Int. J. Accounting Inf. Syst.,* vol. 22, 2016, pp. 44–59.

[21] F. A. Amani and A. M. Fadlalla, "Data mining applications in accounting: A review of the literature and organizing framework," *Int. J. Accounting Inf. Syst.,* vol. 24, 2017, pp. 32–58.

[22] E. W. T. Ngai, Y. Hu, Y. H. Wong, and Y. Chen, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decis. Support Syst.*, vol. 50, no. 3, 2011, pp. 559–569.

[23] G. Özdağoğlu1, A. Özdağoğlu, Y. Gümüş, and G. Kurt-Gümüş, "The application of data mining techniques in manipulated financial statement classification: The case of Turkey," *J. Art. Intell. and Data Mining,* vol. 5, no. 1, 2017, pp. 67–77.

[24] S. Dutta, I. Dutta, and B. Raahemi, "Detecting financial restatements using data mining techniques," *Expert Syst. with Appl.*, vol. 90, 2017, pp. 374–393.

[25] D. Liang, C.- C. Lu, C.- F. Tsai, and G.- A. Shih, "Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study," *Eur. J. Oper. Res.*, vol. 252, no. 2, 2016, pp. 561–572.

[26] P. Ravisankar, V. Ravi, G. Raghava Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," *Decis. Support Syst.*, vol. 50, no. 2, 2011, pp. 491–500.

[27] Y. B. Bhavsar and K. C. Waghmar, "Intrusion detection system using data mining technique: Support vector machine," *Int. J. Emerg. Tech. and Adv. Eng.*, vol. 3, no. 3, 2013, pp. 581–586.

[28] P.- F. Pai, M.- F. Hsu, and M.- C. Wang, "A support vector machine-based model for detecting top management fraud," *Knowl.-Based Syst.*, vol. 24, no. 2, 2011, pp. 314–321.

[29] H. K. Ghritlahre and R. K. Prasad, "Prediction of heat transfer of two different types of roughened solar air heater using artificial neural network technique," *Thermal Sci. and Eng. Prog.,* vol. 8, 2018, pp. 145–153.

[30] W. Huang, K. K. Lai, Y. Nakamori, S. Wang, and L. Yu, "Neural networks in finance and economics forecasting," *Int. J. Inf. Tech. and Decis. Making*, vol. 6, no. 1, 2007, pp. 113–140.

[31] K. Efstathios, S. Charalambos, and M. Yannis, "Audit-firm group appointment: an artificial intelligence approach," *Intell. Syst. in Accounting, Finance and Manage.*, vol. 17, no. 1, 2010, pp. 1–17.

[32] M.- Y. Su, "Real-time anomaly detection systems for denial-of-service attacks by weighted k-nearest-neighbour classifiers," *Expert Syst. with Appl.*, vol. 38, no. 4, 2011, pp. 3492–3498.

[33] C. Sudha and T. N. Raj, "Credit card fraud detection in internet using K-nearest neighbor algorithm," *IPASJ Int. J. Comput. Sci. (IIJCS)*, vol. 5, no. 11, 2017, pp. 22–30.

[34] C. Gaganis, F. Pasiouras, C. Spathis, and C. Zopounidis, "Comparison of nearest neighbours, discriminant and logit models for auditing decisions," *Intell. Syst. in Accounting, Finance and Manage.*, vol. 15, no. 1/2, 2007, pp. 23–40.

[35] Fame, *Our Products*. Accessed on Feb. 5, 2020. [Online]. Available: https://www.bvdinfo.com/en-gb/our-products/data/national/fame#secondaryMenuAnchor2