Artificial Intelligence and Fraud Detection

Yang Bao[1], Gilles Hilary[2], Bin Ke[3]

November 24, 2020

This chapter is forthcoming in Innovative Technology at the interface of Finance and Operations:

Babich V, Birge J, Hilary G (eds) Innovative Technology at the interface of Finance and Operations. Springer Series in Supply Chain Management, forthcoming, Springer Nature

We thank Kai Guo for research assistance.

[1]Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, P.R. China 200030. Tel: +86-21-6293 3672. Email: baoyang@sjtu.edu.cn.

[2] McDonough School of Business, Georgetown University, Washington, D.C., U.S.A. Email: gilles.hilary@georgetown.edu.

[3]Department of Accounting, NUS Business School, National University of Singapore, Mochtar Riady Building, BIZ 1, # 07-53, 15 Kent Ridge Drive, Singapore 119245.Tel: +65 6601 3133. Fax: +65 6773 6493. Email: bizk@nus.edu.sg.

Artificial Intelligence and Fraud Detection

Yang Bao, Gilles Hilary, and Bin Ke

Abstract: Fraud exists in all walks of life and detecting and preventing fraud represents an important research question relevant to many stakeholders in society. With the rise of big data and artificial intelligence, new opportunities have arisen in using advanced machine learning models to detect fraud. This chapter provides a comprehensive overview of the challenges in detecting fraud using machine learning. We use a framework (data, method, and evaluation criterion) to review some of the practical considerations that may affect the implementation of machine-learning models to predict fraud. Then, we review select papers in the academic literature across different disciplines that can help address some of the fraud detection challenges. Finally, we suggest promising future directions for this line of research. As accounting fraud constitutes an important class of fraud, we will discuss all of these issues within the context of accounting fraud detection.

1 Introduction and motivation

Fraud exists in all walks of life. It is an economically significant issue. For example, some studies suggest that losses associated with credit card fraud in the USA alone are close to $17 billion.[1] The recent development in artificial intelligence (AI) in general, and machine learning in particular, has opened potential new venues to tackle fraud. However, recent research by the software firm SAS and the Association of Certified Fraud Examiners suggests that a mere 13% of organizations across industries take advantage of these technologies to detect and deter fraud.[2]

There are different reasons for this relative lack of implementation. Indeed, not all types of fraudulent activities are equally suited for AI treatment. For example, credit card fraud is a good setting for experimenting with machine learning algorithms. The high frequency of credit card transactions provides large datasets required for training, backtesting, and validation of machine learning algorithms. Since a fraudulent activity is rather unambiguously defined, it facilitates the labeling of historical data to train classification algorithms. The historical datasets contain a diversified set of features that can be potentially incorporated in models, ranging from transaction characteristics, cardholder, or transaction history. In contrast, the detection of money laundering activities is more challenging. For example, it is much more difficult to determine if an activity can be legally characterized as money laundering. They may involve multiple parties operating outside the perimeter of the firm. Given the sensitivity of the data involved, financial institutions may be more reluctant to share data (even in pseudo-anonymized format). Unsurprisingly, more progress has been achieved in deploying machine learning techniques to combat credit card fraud than to address money-laundering activities.

[1] https://www.wsj.com/articles/borrower-beware-credit-card-fraud-attempts-rise-during-the-coronavirus-crisis-11590571800

[2] https://www.technologyreview.com/2019/11/18/131912/6-essentials-for-fighting-fraud-with-machine-learning/

Generally, machine learning still struggles with more complex problems. The technology has recently made tremendous progress with facial, voice, and text recognition. In these cases, data is abundant. Some progress has also been achieved in using machine learning to categorize certain events that can be reasonably easily classified by humans. For example, algorithms to detect spam or doctored documents are now reasonably mature. Importantly, as we discuss below, this does not mean that organizations do not face issues when they deploy systems to automate these tasks. In contrast, machine learning may not work well in more complex social situations, especially if they occur over extended periods. In this case, simple rules or human judgment may be more effective. For example, in Salganik et al. (2020), a large group of social scientists (160 teams) tried to predict six life outcomes (e.g., child's grade point average and whether a family would be evicted from their home) using machine-learning methods. The first five waves of data and part of the sixth were available to the researchers; the goal was to predict outcomes in the sixth. The dataset contained close to 13,000 variables for over 4,000 families for 15 years. Data collection included in-depth interviews and in-home observations repeated several times over many years. However, machine-learning tools offered little improvement over standard methods in social science (Garip 2020). Overall, the algorithms barely explained more than a linear regression based on four variables.

We discuss several of these challenges in this chapter. We do not attempt to review the entire literature on fraud, nor do we cover artificial intelligence or even machine learning in-depth. The primary objective of the chapter is to identify the opportunities and challenges associated with deploying machine learning techniques to combat fraud, both from practical and academic perspectives. Many of our statements apply to fraud detection in general. However, to better contextualize our discussion, we will often refer to a specific form of fraud, financial statement manipulation (i.e., accounting fraud).

We focus on financial statement manipulation for several reasons. First, accounting fraud is a worldwide problem. Well-known accounting fraud cases include Parmalat in Europe, Enron in the U.S., and Sino-Forest in China. Second, the consequences of corporate accounting fraud are quite severe. Shareholders, creditors, customers, suppliers, employees, and audit firms all suffer. Karpoff et al. (2008) find that, on average, the fraudulent firms lose 38% of their market values when news of their misconduct is reported.[3] The revelation of accounting fraud cases could also create a significant spillover effect on many innocent companies. For example, Darrough et al. (2020) show that fraud allegations of some Chinese concept stocks result in strong negative spillover effects on the non-fraudulent Chinese concept stocks. Distinguishing between seemingly similar cases is, therefore, important.

Because of the significant costs and externalities associated with fraud, it is important to prevent and detect accounting fraud on a timely basis so that the damages of accounting fraud can be eliminated or mitigated. For example, external auditors are responsible for detecting material fraud in the Statement on Auditing Standards of some jurisdictions.[4] However, as we explain below, accounting fraud detection is a very challenging task for various reasons. Prior research also suggests that existing fraud detection technologies lag, despite the increased frequency and costs of accounting fraud (KPMG 1998; Ernst & Young 2010). However, the significant advances in machine learning and AI technologies offer exciting opportunities to develop more powerful fraud prediction models in recent years.

The rest of the chapter is as follows. In Section 2, we review some of the challenges associated with fraud detection. In Section 3, we use a framework (data, method, and evaluation criterion) to review some of the practical considerations that affect the implementation of

[3] In a 10-year review of corporate accounting fraud commissioned by the Committee of Sponsoring Organization of the Treadway Commission (COSO), Beasley et al. (2010) find that the total cumulative misstatement or misappropriation of nearly $120 billion across 300 fraud cases with available information (mean of nearly $400 million per case) (Beasley et al. 1999).

[4] See SAS no. 99 (AICPA 2002) for a discussion of this issue in a U.S. context.

machine-learning models to predict fraud. In Section 4, we review select papers in the academic literature that can help address some of the challenges discussed in Section 2 and provide a discussion of future directions for this line of research. We conclude in Section 5.

2 Challenges of fraud detection

We start this discussion by reviewing the empirical challenges that are specific to fraud detection using machine learning.

2.1 Problems with fraud and machine learning in general

First, detected fraud cases are rare. For example, the USA's rate of credit card fraud in 2015 was below 0.1%[5]. Most machine-learning algorithms work best when the number of samples in each class is approximately equal because most algorithms are designed to maximize accuracy and reduce error. It is hard for algorithms to learn when samples are really unbalanced as they do not frequently encounter fraud cases. The existing literature is aware of this problem and has resorted to various methods to deal with this challenge.

Second, fraud is adversarial. Machine-learning techniques work best when patterns are stable, and important data are not omitted systematically, or worse, manipulated. In many applications of machine learning, parties are cooperating with the system to facilitate its learning (e.g., medical applications), or at worst, are neutral toward it. In the case of fraud, perpetrators try to prevent the learning. For example, fraudsters may open accounts in different financial institutions in different jurisdictions to prevent an effective network analysis. Further, fraudsters are constantly imagining new schemes. Thus, there may not be a precedent for algorithms to detect a new type of fraud.

---

[5] https://www.federalreserve.gov/publications/files/changes-in-us-payments-fraud-from-2012-to-2016-20181016.pdf

2.2 Problems that are specific to accounting fraud and machine learning

Aside from these general concerns, specific types of fraud can have specific issues. We focus on accounting fraud to present examples of domain-specific issues.

First, many accounting fraud cases remain undetected or at least take a long time to be recognized. Most credit card frauds can be quickly identified, and a determination of the existence of a fraud after a complaint is straightforward in most cases. In contrast, determining the existence of an accounting fraud requires significant investigative resources in many instances, and the distinction between creative but legitimate practices and fraud may be difficult to ascertain. Models that explicitly consider the fact that fraud may be undetected are rare. Further, accounting fraud cases are often detected by regulators. In contrast, other types of fraud may be identified by the victim. To the extent that regulators are ineffective or biased, models of fraud detection will be ineffective and biased. For example, Dyck et al. (2020) estimate that only about half of severe financial reporting violation cases are detected by the Security Exchange Commission (SEC). This issue is likely to be even more serious in less developed countries with weaker institutional environments.

Second, if a firm commits accounting fraud, it tends to misstate accounting reports for several years before being caught. Most existing fraud prediction studies do not consider this serial fraud feature in model building. Instead, they tend to treat each firm-year as an independent observation, ignoring the time series dependence of serial fraud cases. Studies that consider the serial fraud issue sidestep this problem by using only the initial fraudulent year (e.g., Amiram et al. 2015) or remove the serial fraud observations in the training year or test year (e.g., Brown et al. 2020).

Lastly, due to the time variation in managerial incentives and monitoring intensity, accounting fraud behavior exhibits regime shifts both over time and cross-sectionally (Beasley et al. 1999; Beasley et al. 2010). Prior research (Dechow et al. 2011) shows that the frequency of

accounting fraud varies significantly across industries. Except for Abbasi et al. (2012), the existing accounting fraud prediction models have not incorporated such regime shifts in model building.

3 Practical considerations in model building

As noted by Abbasi et al. (2012), building any fraud prediction model requires a researcher to make important decisions on the following three crucial components: (i) What data inputs (predictors) and data outputs (fraud labels) should be used for the model? (ii) What specific machine learning methods should be used for the prediction task? (iii) What evaluation criteria should be used to judge the performance of a fraud prediction model? Below we elaborate on each of the above three components of model building.

3.1 Data

3.1.1 Data and Fraud

Data quality is crucial for AI to be effective. As systems go from rules to simple structural models, from structural models to machine learning, from machine learning to deep learning, the amount and the quality of data that models require increase. This creates several challenges.

Quantity

The first one is the quantity of data. This issue is progressively solved as the cost of acquiring and storing structured data in data warehouses and unstructured data in data lakes decreases. However, for smaller organizations, infrequent situations, or extremely complex estimations, this can remain an issue.

Quality

A second issue is the quality, integrity, and comprehensiveness of data. A torrential flow of data emerging from heterogeneous sources has no utility if it is not structured properly.[6] Data maintenance alone is a challenge. Indeed, many financial institutions have to contact thousands of customers every month to refresh Know-Your-Customer (KYC) documents and update information that is incorrect or missing in their databases. This issue is a particularly vexing problem if the validation must happen in continuous time. In addition, dealing with a legacy system can be difficult. A good example is the Michigan Unemployment Insurance Agency's (UIA) switch from a 30-year-old mainframe system running COBOL to a new system dubbed Michigan Integrated Data Automated System (MiDAS). After spending over $44 million and 26 months on the project, the UIA launched MiDAS in October 2013. Soon after, the number of persons suspected of unemployment fraud grew fivefold compared to the average number found using the old system. Using MiDAS generated savings of close to $70 million for the agency.[7] Unfortunately, the error rate of the new system was staggering (around 90%). In August 2015, the agency stopped using the system without human intervention, but by then, MiDAS may have already falsely accused at least 20,000 people. The problem was at least partly rooted in the transfer of scanned documents to the new system.[8]

More generally, legacy systems typically suffer from integration problems when they need to interact with machine learning platforms. Their workflow, data management, and change controls are often poorly aligned with the needs of a modern machine learning system. This problem is mitigated when the machine learning system is deployed in parallel (e.g., an ex-post analysis of transactions to meet anti-money laundering requirements) but exacerbated when directly integrated into the workflow (e.g., in a payment process).

---

[6] See Zhang, Yang, and Appelbaum (2015) for a good discussion of these issues.

[7] https://spectrum.ieee.org/riskfactor/computing/software/michigans-midas-unemployment-system-algorithm-alchemy-that-created-lead-not-gold

[8] https://www.freep.com/story/news/local/michigan/2017/07/30/fraud-charges-unemployment-jobless-claimants/516332001/

One may assess the quality of big data along four dimensions: volume, velocity, variety, and veracity. Unfortunately, the first three dimensions can impede the last. For example, heterogeneous data formats can affect data consistency. Firms normally decide what data to collect primarily based on legal and operational grounds. It is common for different operational entities within the same organization to develop their own policies for capturing, cleansing, normalizing, storing, searching, sharing, analyzing, and retaining data. Datasets are often segregated logically and physically across departments and geographical regions. Legal considerations may prevent the smooth integration of these data sets. For example, laws protecting national security typically prohibit the export of any data that falls within their scope, even if the data is shared within the same organization. China's state secret laws, which make the export of any state secret a criminal act, are examples of this type of regulation. Naturally, sharing across organizations is even more complicated. For example, the UK Data Protection Act allows for the storage and exchange of data between organizations. If a credit fraud-checking agency detects an anomaly between a new application for a lender and previous applications for other lending institutions, the agency can make each institution aware of the concern. In contrast, Ireland has not authorized this type of information sharing.[9]

Even absent legal issues, data integration can be difficult for technical reasons. The sheer size and complexity of data sets can make them difficult to manage and process using available database management tools. Nonstandard data structures, overlapping production cycles, and fragmented sources make data aggregation across legal entities, subsidiaries, and vendors difficult. Furthermore, data encryption and data stored on blockchains pause additional challenges.

Data that need to be moved and shared across departments or locations are generally provided in a summarized format and are limited to supporting specific functions, such as

[9] http://www.eurofinas.org/uploads/documents/Non-visible/Eurofinas-Accis_ReportOnFraud_WEB.pdf

finance (e.g., invoicing, payment) or operations (e.g., shipping). These records are not necessarily the most relevant for fraud detection. For example, transactional data (e.g., phone call records) may not be readily available across departments.

Data mapping and traceability become key factors, and computer-assisted solutions have been designed to facilitate these processes, but naturally, they have their limitations. For example, while most enterprise resource planning (ERP) systems have certain fraud prevention and detection capabilities, systems often turn off controls to function more efficiently.[10]

Excess of data

A third issue is (somewhat paradoxically) an excess of data. Over time, a legal structure has been developed to regulate the collection and storage of information. For example, the GDPR (General Data Protection Regulation) is a European data regulation that carries a maximum fine of the greater of €20 million or 4% of annual global turnover for non-compliance. One of its provisions is that organizations are responsible for securing the data they collect. The British data regulator, Information Commissioner's Office (ICO), proposed a £183m fine in 2019 for a 2018 breach. In the USA, a patchwork of state regulations has also gradually increased the disclosure requirements (see Chen, Hilary, and Tian (2020) for an analysis of these laws). Aside from protecting the data once they have been collected, the organization needs first to pay attention to the legality of the data collection. For example, the Internet regulator in China investigated smartphone applications to determine if they collect users' information illegally or excessively. The Ministry Industry and Information Technology targeted several popular applications such as Tencent's QQ and QQ Reading, Xiaomi's digital finance app Xiaomi Finance, and the inter-city delivery service FlashEX in 2019.[11] In this context, the ever-increasing collection of data to feed the algorithms is not necessarily optimal. A standard

[10] https://www.corporatecomplianceinsights.com/the-growing-problem-of-corporate-fraud/
[11] https://technode.com/2019/12/19/tencent-xiaomi-apps-called-out-for-illegal-data-collection/

cost-benefit analysis should be applied to data collection and storage and reflect the regulatory and reputation risk in case of breaches or misuses.

Biases

Lastly, biases can be a significant issue.[12] Suggest re-write sentence as "Labeled training data can be biased because individuals have treated them with implicit or explicit biases. Ethnic biases have been well-publicized. For example, a recent NIST study reports the presence of demographic effects in U.S. face recognition algorithms, but due to biases in the type of photos used.[13] Employing socially diverse individuals may mitigate this issue.

However, biases may also be introduced into algorithms if they learn from a third-party dataset contaminated by bias. For example, the Equal Credit Opportunity Act (ECOA) prohibits credit discrimination based on race, color, religion, national origin, sex, marital status, age, or because a person receives public assistance. If a company created a score to make credit decisions based on consumers' Zip Codes, resulting in a "disparate impact" on particular ethnic groups, this would likely be an illegal practice under ECOA.[14] This algorithm would also be contaminated if this score was subsequently incorporated in a dataset used in fraud detection.

However, the issue of biases is broader than the illegal ones. For example, if an algorithm analyzes official communication in a bank, there may be a tendency for traders who want to engage in nefarious activities to communicate through private channels. This would then create potential biases in the analysis by systematically omitting important information.

3.1.2 Data and financial statement fraud

[12] Supervised models "learn" from labeled data. To train a supervised model, one presents both fraudulent and non-fraudulent records that have been labeled as such. Unsupervised models ask the model to "learn" the data structure on its own.

[13] https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf

[14] https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms.

Most financial statement fraud prediction studies employ supervised learning that requires data on both fraud labels (the dependent variable) and fraud predictors (the independent variables). The first critical decision one needs to make in building a fraud prediction model is to select an appropriate accounting fraud database. There are a variety of input data one could use to build fraud prediction models. Many studies utilize one data source only. For example, Dechow et al. (2011) focus on readily available accounting data, Cecchini et al. (2010) and Purda and Skillicorn (2015) use textual data only, while Dong et al. (2018) combine textual and network data together. However, few studies have attempted to combine all available input data of different kinds in building a unifying fraud prediction model.

Partially reflecting the ambiguity about what constitutes fraud, prior research has used different terms to describe fraud or alleged fraud, such as fraud, misconduct, irregularities, misreporting, and misrepresentation. Accordingly, the existing accounting fraud literature has used various databases to measure fraud, including the Government Accountability Office (GAO), Audit Analytics databases of restatement announcements, the Stanford Securities Class Action Clearinghouse (SCAC) database of securities class action lawsuits, and the SEC Accounting and Auditing Enforcement Releases. Karpoff et al. (2017) provide a detailed comparison of the pros and cons of using the four different databases for financial misconduct research. They find that the results from empirical tests can depend on which database is accessed and offer suggestions for researchers using these databases.

The second critical decision one has to make in building fraud prediction models is to select the list of fraud predictors. With the explosion of big data in the past decade, there exists a variety of predictors one could use to build accounting fraud prediction models. One may be tempted to throw in as many predictors as possible into a machine learning model for training and prediction, but Bao et al. (2020) confirm that more predictors do not necessarily improve prediction performance. Also, more costs are incurred when using more data, and the prediction

model would be less generalizable to different countries and industries. Hence, a cost-benefit framework is also warranted in selecting the fraud predictors.

Following the cost-benefit framework, we could classify the fraud predictors based on the nature of the input data: structured data (e.g., accounting numbers) versus unstructured data (e.g., text, video, and voice). Unstructured data such as text or video are much harder to process than structured data and hence more costly. In addition, the useful information embedded in unstructured data could be already contained in the structured data. Given the increasing availability of many structured data sources, it makes sense to first extract as much useful information from structured data sources as possible before turning to the unstructured data sources for fraud prediction.

One could also classify the fraud predictors based on the causal motivation-ability-opportunity framework from the criminology literature. If one believes that the motivation-ability-opportunity framework is relatively comprehensive in explaining accounting fraud behavior, one could build powerful fraud prediction models based on theory-motivated input data as predictors. Past studies have developed many proxies for motivation (e.g., Burns and Kedia 2006) and ability/personal traits (e.g., Davidson et al. 2015) while others have developed proxies for opportunity (e.g., Beasley 1996; Larcker et al. 2007). Hence, researchers could select such theory-motivated fraud predictors in model building.

3.2. Methods

As analysts and researchers typically code fraud into a binary variable, logistic regressions have traditionally been the most popular learning method in the business literature before the rise of the machine learning field (e.g., Dechow et al. 2011). With the increasing availability of many unstructured databases and the emergence of methodological breakthroughs, researchers and analysts start to employ more sophisticated learning methods to train and predict

accounting fraud. For example, Cecchini et al. (2010) and Perols et al. (2017) use support vector machines (SVM) to train a fraud prediction model. Amiram et al. (2015) apply Benford's Law to fraud prediction.

3.3 Evaluation metrics

3.3.1 Evaluation metrics in general

One issue with algorithms is what they should be maximizing. Often, the objective is to maximize accuracy under the assumption that efficient algorithms can both minimize false positives and false negatives. However, the two categories of errors need not entail similar private or social costs. More complex loss functions can be designed and incorporated into the algorithms. These different trade-offs need to be analyzed before algorithms are defined.

However, the definition of the loss function is more than a technical one. For example, organizations need to balance detection and customer experience. An algorithm that provides better classifications but requires intrusive data requests may be sub-optimal. This concern becomes more relevant as regulation imposes more constraints on what kind of data organizations can collect, and future regulations may also impose constraints on the algorithm loss function. This potential disconnection between algorithm design and production constraints may explain why some statistics suggest that only 50% of all models developed ever make it into production.[15]

3.3.2 Evaluation metrics for accounting fraud prediction models [16]

[15] https://mit-insights.ai/6-essentials-for-fighting-fraud-with-machine-learning/
[16] This section heavily relies on Bao et al. (2020).

There are different ways to assess the performance of prediction models. The first classification performance metric considered is *accuracy*, defined as $\frac{TP+T}{TP+FN+FP+T}$, where TP (true positive) is the number of fraudulent firm-years that are correctly classified as fraud; FN (false negative) is the number of fraudulent firm-years that are misclassified as non-fraud; TN (true negative) is the number of non-fraudulent firm-years that are correctly classified as non-fraud; and FP (false positive) is the number of non-fraudulent firm-years that are misclassified as fraud. Bao et al. (2020) rejected *accuracy* as appropriate due to the imbalanced nature of our fraud versus non-fraud data. Over the period 1979–2014, the frequency of fraud detected by U.S. regulators is very low, typically less than 1% of all firms per year. Hence, a naïve strategy of classifying all firm-years as non-fraud in their sample would lead to an accuracy of better than 99% based on *accuracy*. However, such seemingly high-performance fraud prediction models are of little value in our fraud prediction task because we care about both the true negative rate (i.e., *specificity*) and the true positive rate (i.e., *sensitivity*).

To properly gauge the performance of a fraud prediction model, Bao et al. (2020) also considered but rejected balanced accuracy (*BAC*) as an alternative performance evaluation metric (He and Ma 2013). *BAC* is defined as the average of the fraud prediction accuracy within fraudulent observations and the non-fraud prediction accuracy within non-fraudulent observations. Specifically, $BAC = \frac{1}{2} * (Sensitivity + Specificity)$, where $Sensitivity = \frac{TP}{TP+}$ and $Specificity = \frac{TN}{TN+F}$. Larcker and Zakolyukina (2012) note two important limitations of *BAC* as a performance evaluation metric. First, *BAC* is constructed based on a specific predicted fraud probability threshold of a given classifier, and the threshold is usually automatically determined by the classifier to maximize the *BAC*. In the absence of any knowledge of the costs of misclassifying false positives versus the costs of misclassifying false negatives, one cannot determine the optimal predicted fraud probability threshold for the purposes of classifying

fraud and non-fraud. Second, measures such as S*ensitivity* are very sensitive to the relative frequency of positive and negative instances in the sample (i.e., data imbalance).

To avoid *BAC*'s limitations, Bao et al. (2020) adopt *AUC* as one performance evaluation metric, following Larcker and Zakolyukina (2012). *AUC* is the area under the Receiver Operating Characteristics (ROC) curve. A ROC curve is a two-dimensional depiction of a classifier's performance that combines the true positive rate (i.e., *sensitivity*) and the false positive rate (i.e., 1-*specificity*) in one graph (Fawcett 2006). *BAC* represents only one point in the ROC curve. Many fraud prediction models use *AUC* as the primary performance evaluation metric.

Recall that the average frequency of detected accounting fraud among publicly listed U.S. firms is less than 1%. Therefore, even a top performing fraud prediction model (e.g., Cecchini et al. 2010) would generate a large number of false positives. Table 7 in Cecchini et al. (2010) illustrates this point: their SVM with a financial kernel correctly classifies 80% of the fraud observations and 90.6% of the non-fraudulent observations in the out-of-sample test period, the best among the competing models considered in their study. However, applying the Cecchini et al. model to the test period 2003-2008 considered by Bao et al. (2020) would result in too many false positives. Specifically, fraud occurred in only 237 of the 30,883 firm years during the test period 2003-2008. Cecchini et al.'s method, however, would mislabel 2,881 ((1-90.6%)*(30,883 - 237)) non-fraudulent observations as fraud—a serious overestimate of the number of actual cases of fraud in the test period.

To deal with this problem, Bao et al. (2020) introduce a new performance evaluation metric to the fraud prediction literature by treating the fraud prediction task as a ranking problem. Specifically, we can limit the out-of-sample performance evaluation to only a small number of firm years with the highest predicted probability of fraud. In this scenario, the performance of a fraud prediction model can be measured by the following performance evaluation metric for ranking problems: Normalized Discounted Cumulative Gain at the position *k*

(NDCG@k). NDCG@k is a widely used metric for evaluating ranking algorithms such as web search engine algorithms and recommendation algorithms (Järvelin and Kekäläinen 2002) and has been theoretically proven effective (Wang, Wang, Li, He, Chen, and Liu 2013). The values of NDCG@k are bounded between 0 and 1.0, and a higher value represents a model's better-ranking performance. NDCG@k avoids the aforementioned problem of investigating too many false-positive cases by limiting the investigation to no more than a given number *k* of firm years with the highest predicted fraud probability in the test period. As the average frequency of detected accounting fraud among publicly listed U.S. firms is less than 1%, Bao et al. (2020) set *k* equal to the top 1% of the firm years in the test period.

3.4 Caveats about ML models

Machine-learning can be useful in the context of combating fraud in three ways: detection & interdiction, litigation, and prevention. An example of detection and interdiction is when a financial institution blocks credit card transactions in real time in case of suspicious activity. Litigation is a situation in which machine learning analysis is used to build a legal case. Prevention is an approach in which an organization uses machine learning insights to conduct a root cause analysis, reorganize its operations, and minimize the fraud risk in the first place. Each approach comes with its own set of issues.

In the case of detection and interdiction, the output of many algorithms is a score that predicts the risk level associated with a situation. However, users have to be clear about what is being predicted. In many cases, the algorithm will detect anomalous transactions (i.e., those that are different from an expected benchmark), but an anomalous transaction does not necessarily mean a problematic transaction. If the analysis is not conditioned properly, the lack of apparent anomaly can be problematic. For example, Parmalat (the largest European accounting fraud case to date) had very stable accounting ratios before the fraud was revealed.

Further, what is normal can be unstable. For example, changes in the regulatory environment or socio-economic environment (e.g., Covid 19) can modify people's behavior and require a new benchmarking.[17] Naturally, algorithms can learn, but the adjustment is unlikely to be immediate. Also, many algorithms are based on non-linear statistics that yield unstable models. For example, several studies have shown that including an incongruous element (e.g., the picture of an elephant) can lead neural networks that otherwise detect and categorize objects (e.g., a person, a couch, a television) with high confidence to miscategorize the objects completely.[18]

Another issue with fraud prediction score is explainability. Current machine learning algorithms are notoriously bad at dealing with this issue. The importance of this weakness varies with the objective of fraud detection. If the objective is prevention through interdiction (e.g., credit card), the need for explainability can be lower. However, even in this situation, the issue cannot be ignored. For example, the European GDPR requires "the data subject shall have the right to obtain from the controller […] information [about] the existence of automated decision-making, including profiling, […] and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject." Human intervention is key, and processes must allow for a review by someone who has the appropriate authority and capability to change the decision generated by the system. An overtly complex algorithm may hinder compliance in that respect.

If the objective is to prevent fraud through litigation, the motivation of the classification becomes central, and there is a need for due process in an electronic context (e.g., Citron 2008). It is also important in these cases to distinguish between aggressive (but legal) and fraudulent (and illegal cases) situations. While the difference can be clear-cut in some cases (e.g., the card

[17] https://www.technologyreview.com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud-humans-in-the-loop/

[18] https://arxiv.org/abs/1808.03305

owner authorized the payment or not), the difference between the two can be difficult to elucidate in complex cases (e.g., accounting fraud). Although machine learning can also conceivably help tackle this question, this analysis introduces an additional layer of complexity in the classification.

If the objective is to prevent fraud through anticipation (e.g., process reengineering), the motivation for the classification has to be deeply understood so that a root cause analysis can be conducted and remedial actions implemented. This process may involve analyzing operational data (e.g., internal incentives) that are typically not included in many current fraud detection algorithms. Further, going from a cluster of suspect cases generated through machine learning to a change in processes and organizations may be difficult. For example, it is well known that humans in general, and auditors in particular, have limited ability to process large amounts of information required for complex decision making (e.g., Iselin 1988). Prior research (e.g., Ashton 1974) has shown that large volumes of accounting information can lead to suboptimal financial and auditing judgments. In this context, machine-learning algorithms can act as data reduction tools to economize mental resources. However, the ability of individuals to combine cues from multiple sources is also limited (e.g., Benbasat and Taylor 1982). The cognitive integration of black box results from algorithms with complex organizational structures is poorly understood currently, but it seems plausible that this integration may bring new challenges.

3.5 Distinction between prediction and causal inference[19]

Causal inference and prediction are fundamentally different problems. The objective of causal inference is to use statistical tools to test causal relationships. In contrast, the objective of prediction is to apply a statistical model or data mining algorithm to data for the purpose of

[19] This section borrows heavily from the online appendix of Bao et al. (2020).

predicting *new observations* (Hastie, Tibshirani, and Friedman 2009; Shmueli 2010; Kleinberg et al. 2015).

The distinction between prediction and causal inference is significant because most existing accounting and finance academic research focuses on causal inference, while practitioners may be more interested in prediction. Causal inference studies are relevant for decision-makers who wish to design effective policy remedies to prevent and mitigate accounting fraud. However, there are many important decisions (e.g., whether or not to invest in a high-growth stock) that require an accurate and timely prediction of whether a firm is or is not engaged in fraud (i.e., a prediction problem). Furthermore, causal inference and prediction are not mutually exclusive. For example, Varian (2014) argues that predictive modeling can also benefit causal inference research because predictive modeling can provide a low-cost estimate of unobservable counterfactuals in causal inference (i.e., the outcome that would have happened without the policy intervention). Varian (2014) even argues that a good predictive model can be better than a randomly chosen control group due to imperfection of the randomization process.

The distinction between prediction and causal inference has a few important implications. First, causal inference emphasizes the unbiasedness of regression coefficients, but prediction may deliberately increase the bias of a regression coefficient in order to minimize the out-of-sample prediction error. Hence, for prediction purposes, a "wrong" model could produce better performance than a correctly specified model. Second, while causal modelling requires that f represent an underlying causal function, predictive modelling requires only an association between x and y. That is, an input variable that is not causal (e.g., raw financial data items that may have no obvious economic interpretation) could be included in a prediction model. Third, the choice of f could be different for causal inference and prediction problems. While f is carefully chosen based on theory and causal relationship for causal inference problems, f is often constructed from data and could take on more flexible and complex functional forms.

4 A brief overview of existing academic research on fraud detection

Fraud prediction is highly interdisciplinary literature. We divide this literature into two streams by discipline: (i) studies that focus on fraud prediction out of sample using machine-learning methods in the non-accounting academic fields (e.g., computer science) and (ii) studies in the accounting academic literature that focus on causal inference. Because these two types of literature often use different methodologies and emphasize different aspects (causal inference vs. prediction), we review the two literature streams separately. We adopt the conceptual framework in section 3 (data, method, and evaluation criterion) to organize the literature overview.

4.1 Fraud prediction using machine learning in the non-accounting academic fields

Sections 2 and 3 have outlined some of the challenges associated with fraud detection through machine learning. The machine learning literature in the non-accounting fields has proposed various solutions to some of the aforementioned challenges. These models attempt to predict different types of fraud, such as credit card fraud, insurance fraud, e-commerce fraud. Interestingly, we find no study in the machine learning academic literature outside the accounting field that examines accounting fraud detection. The building of these models usually follows the similar conventional supervised machine learning pipeline. Specifically, fraud detection is treated as a binary classification problem. The prediction model is trained using off-the-shelf supervised machine learning algorithms, including but not limited to Logistic Regression, Support Vector Machine (SVM), Decision Tree, Neural Network, Random Forests. Since these conventional models have already been summarized in previous survey papers (Ngai et al. 2011; Dutta et al. 2017; Hajek and Henriques 2017), we only highlight some recent fraud detection models in the non-accounting fields that have significant methodological contributions addressing the challenges of fraud detection.

First, as noted in Section 2, detected fraud is typically a rare event. One commonly used method to deal with rarity is to reduce the imbalance of fraud and non-fraud observations by matching (e.g., Green and Choi 1997; Lin et al. 2003; Whiting et al. 2012; Fletcher, Glancy and Yadav 2011; Humpherys et al. 2011; Cecchini et al. 2010). While it is appropriate to use a smaller matched sample of fraud and non-fraud for model training, it is problematic to use only the matched fraudulent and non-fraudulent firm-years in the holdout test sample to evaluate the out-of-sample performance of a prediction model. This is because doing so would invite look-ahead bias.

Second, fraud detection is difficult because many fraud cases remain hidden. While the rare fraud case issue can be addressed reasonably well by combining imbalance learning and ensemble learning, the hidden fraud issue is seldom considered. To handle these two issues, Li et al. (2014) propose to detect fake reviews by using positive-unlabeled (PU) learning (Bekker and Davis 2020). PU learning is a naturally suitable method for addressing the two issues simultaneously because it can be used to learn a binary classifier by only using positive examples (i.e., true fraud cases) and unlabeled examples (i.e., unknown cases that could be either fraud or non-fraud). In addition to PU learning, the other possible methods are one-class classification, unsupervised learning, transfer learning, and adversarial learning. For example, de Roux et al. (2018) propose an unsupervised machine learning model for detecting fraudulent taxpayers without using any labeled data. Zheng et al. (2019) propose a one-class classification model based on generative adversarial networks that use only one class of samples (i.e., fraud cases) for fraud detection. Zhu et al. (2020) propose a transfer learning framework for cross-domain fraud detection by transferring knowledge from existing domains with abundant labeled fraud cases to improve the new domain's performance with rare labeled fraud cases. Fiore et al. (2019) propose improving credit card fraud detection by using generative adversarial networks for generating minority class examples (i.e., pseudo fraud cases).

Third, serial fraud, where a fraudster commits fraud in several consecutive periods before being caught, is quite common due to delay in detection. Almost all existing models treat each fraud-period as an independent fraud incidence and ignore the time series dependence of serial fraud cases. To address this issue, Guo et al. (2018) adapt the commonly used recurrent neural network model LSTM (Long Short-Term Memory) to extract the useful features from serial fraud behaviors. Oentaryo et al. (2014) also find that the extracted time-series features can improve fraud detection performance.

Fourth, fraud evolves over time, and fraudsters are highly adaptive because they learn from the detected fraud cases. To build adaptive fraud detection models, Abbasi et al. (2012) propose a meta-learning framework that can be learned in a self-adaptive manner to improve prediction accuracy. Xu et al. (2017) propose an online learning algorithm for reputation fraud campaign detection, which can be efficiently updated based on the new fraud cases for adaptively capturing the regime shift.

Last but not least, there is a growing body of research on using multimodal machine learning (Baltrušaitis et al. 2018) for improving fraud detection. In addition to the accounting data, there are various heterogeneous data types (e.g., text, image, audio, video, network) that could contain useful information for fraud detection. Representative examples include text-based fraud detection (Cecchini et al. 2010; Humpherys et al. 2011; Purda and Skillicorn 2015; Wang and Xu 2018) and network-based fraud detection (Beutel et al. 2015; Van Vlasselaer et al. 2016; Shah et al. 2017; Yuan et al. 2017; Dong et al. 2018; Cao et al. 2019; Hu et al. 2019; Liang et al. 2019; Liu et al. 2019; J. Wang et al. 2019; D. Wang et al. 2019; Zhong et al. 2020).

4.2 Fraud prediction in the accounting literature

There is a long literature in accounting on fraud prediction. However, most accounting fraud prediction studies deal with causal inference (i.e., what factors causally affect the incidence of

accounting fraud) rather than fraud prediction. Even if some studies use the term "fraud prediction," they often mean fraud prediction *in sample* rather than out of sample (e.g., Brazel et al. 2009; Hobson et al. 2012).

Because many fraud prediction studies in the accounting field deal with causal inference, we adopt the well-known motivation-ability-opportunity framework from criminology to organize our review. The criminology framework considers three crucial factors in explaining an individual's criminal activities: (i) whether the person has the motive to commit a crime, referred to as incentive variables; (ii) whether the person has the ability to commit a crime (e.g., does the person have a gun?); and (iii) whether the person has the opportunity to commit a crime (e.g., whether the person is at the crime scene).

One could measure any of the above three classes of factors using non-accounting data. For example, one could measure a person's incentive to commit fraud using the person's compensation contract details. As accounting data reflects a firm's business activities, one could also use accounting data to construct indirect proxies for the three classes of factors. For example, one could test whether growth firms are more likely to commit fraud by using a firm's sales growth or market-to-book ratio to measure firm growth.

The existing business literature has considered all three classes of factors to explain firms' accounting fraud behavior. In terms of motivation factors, prior research has considered the effects of both capital markets (e.g., incentives to increase a firm's stock price) and contracts (e.g., incentives to mitigate the constraints imposed by explicit business contracts). Representative research in the first category includes Dechow et al. (1996), Beneish (1999), Burns and Kedia (2006), Efendi et al. (2007), and Johnson et al. (2009). Representative research in the second category includes Healy (1985), Dechow et al. (1995), Dechow et al. (1996), and Burns and Kedia (2006). In terms of ability factors, representative research includes Beneish (1997), who shows that high lagged accruals can help identify earnings manipulation firms.

High lagged accruals are thought to indicate that managers have exhausted legitimate techniques for earnings management. In terms of opportunity factors, prior research has considered the role of corporate governance variables in explaining accounting fraud. Representative research includes Beasley (1996), Dechow et al. (1996), and Larcker et al. (2007).

Much existing accounting research uses readily available financial statement data (e.g., Dechow et al. 2011). Recent years have also witnessed accounting researchers' increasing usage of textual data (e.g., Cecchini et al. 2010; Larcker and Zakolyukina 2012; Purda and Skillicorn 2015; Brown et al. 2020) and network data sources for fraud prediction (e.g., Dong et al. 2018).

As accounting fraud is treated as a binary variable in most prior accounting research and most studies in this field focus on causal inference, logistic regressions are the most common statistical method in explaining the determinants of accounting fraud. True out-of-sample prediction studies are still relatively rare in the accounting literature. However, there has been a surging interest among accounting researchers in using interdisciplinary methods to predict accounting fraud out of sample. Representative research includes Cecchini et al. (2010), Larcker and Zakolyukina (2012), Abbasi et al. (2012), Purda and Skillicorn (2015), Perols et al. (2017), Dong et al. (2018), Brown et al. (2020), and Bao et al. (2020).

4.3 Future research directions

We believe that it would be a fruitful venue for future research to combine the knowledge and expertise from both the accounting and machine learning domains to develop more powerful and adaptive learning models. In particular, we believe that the following interdisciplinary challenges deserve special attention by researchers interested in such problems.

First, it is important to incorporate regime shift in prediction models. Existing research in other disciplines has already developed various methods based on adaptive learning (e.g.,

Brazdil et al. 2008; Abbasi et al. 2012). Another possible approach is to employ "online" learning algorithms to dynamically adapt to the recent new fraud patterns (Xu et al. 2017). Online learning is an important family of machine learning algorithms in which data arrive in a sequential order and the model is updated incrementally (Hoi et al. 2018). This approach is different from the traditional "offline" machine learning algorithms that learn on the entire training data at once.

Second, undetected accounting fraud is a worldwide problem, especially in emerging markets. Hence, future researchers must develop models that can uncover such undetected fraud cases so that relevant decision-makers can take necessary intervention actions to deter and uncover fraudulent behavior. Existing research usually ignores this challenging problem and simply treats undetected accounting fraud cases as clean non-fraud cases. The possible direction for future research is to use unsupervised, semi-supervised, and positive-unlabeled learning algorithms (Bekker and Davis 2020) to learn fraud detection models from truly clean fraud and non-fraud cases.

Third, future researchers could also explore the possibility of building more powerful models by infusing causal theories into machine learning. As noted in Section 4.2, the accounting literature has identified many causal determinants of accounting fraud. For example, Beasley (1996) finds that firms with larger proportions of outside members on the board of directors are less likely to commit accounting fraud. While existing research has attempted to use causal theories in selecting the fraud prediction models' input data, we are not aware of any existing study that incorporates the direction of a causal relation between an input variable and the output variable (i.e., accounting fraud) in building machine learning models to predict accounting fraud.

Fourth, considering the fact that serial fraud is prevalent in reality, future researchers could consider whether it is possible to build more accurate fraud prediction models by

considering the time-series dependence of accounting fraud in model building. It seems promising to manually construct time-series of fraud features and feed them into conventional fraud prediction models (Oentaryo et al. 2014) or use time-series models such as recurrent neural networks for modeling serial fraud behavior directly (Guo et al. 2018).[20]

Fifth, publicly listed firms are required to produce many reports. The easy-to-process financial statements alone contain hundreds of accounting accounts. Existing research has only utilized a small portion of these data. Hence, it is potentially fruitful to employ powerful machine learning models such as deep learning to extract more useful information from such raw accounting data for fraud prediction (Zhang et al. 2019). It is also an interesting research direction to learn better fraud detection models from multimodal data, including but not limited to multilingual texts, images, audios, videos, and networks.

As we have illustrated in the previous sections, modeling fraud prediction is an interdisciplinary task that requires the close collaboration of accounting domain experts and machine learning experts. Hence, we strongly encourage experts across disciplines to come together to build better quality fraud prediction models that can be readily adopted by companies. We strongly believe that such interdisciplinary collaboration will have a better potential to generate breakthrough fraud prediction models.

As noted in Section 3, it is not a trivial task to define accounting fraud and assemble a large sample of accounting fraud. Hence, we also encourage the fraud prediction research community to work together to build a common and better accounting fraud database so that future researchers do not have to repeat the dirty work of data collection and cleaning datasets repeatedly. By having a common database, future research can more easily compare the different fraud prediction models' performance.

[20] Recurrent neural networks are artificial neural networks where connections between nodes form a directed graph along a temporal sequence.

5. Conclusion

Deploying machine learning to detect anomalies, errors, and fraud is a promising and growing research field. In this chapter, we discuss the most critical challenges in fraud detection and highlight many important empirical considerations in fraud prediction model building. We also provide a broad overview of the state-of-the-art approaches to predicting fraud in the extant literature and suggest promising future research directions.

The ideas and algorithms discussed in this chapter should be of interest to academic researchers and decision-makers in many organizations. For example, machine-learning platforms may allow organizations to monitor transactions in nearly real-time. These platforms may allow for a comprehensive analysis (rather than sampling) and faster remediation. Textual analysis can be delegated to a greater extent. For example, machine learning platforms can already read and analyze complex lease agreements to determine their appropriate accounting classification. In turn, this will allow for cleaner and faster closings of the books while providing a better audit trail to detect internal anomalies. Textual analysis can identify relevant documents through terabytes of data through Topic segmentation and keyword analysis. For example, the British Serious Fraud Office (SFO) exposed large-scale bribery and corruption at Rolls-Royce. They used machine learning to sift through 30 million documents, processing up to 600,000 every day. In contrast, the Chief Technology Officer (CTO) of the SFO indicated that the average processing rate of lawyers is 300 documents a day with lower accuracy and consistency.[21] These documents can also be automatically translated and summarized. Sentiment analysis can, for example, allow for the detection of stress, a predictor of fraud. Internal networks of communication can be easily mapped. As new data sources emerge (e.g., Internet of Things, drones), new internal controls can be designed. For example, one can envision that

[21] https://www.cio.com/article/3525877/serious-fraud-office-cto-ben-denison-reveals-how-ai-is-transforming-legal-work.html

automatic treatment of videos could allow for the continuous physical monitoring of inventory and their reconciliation with accounting records. External sources of data can also be integrated. For example, video and images of physical assets from Google Earth and other geographic information system applications can help assert the plausibility of different assumptions. Social network analysis can allow for the anticipation of product issues.

However, the deployment of these tools in organizations remains challenging. Aside from the technical issues outlined in this chapter, these tools will challenge existing practices and require change management skills. As fraud detection goes from an experience-based approach to a data-driven approach, the articulation of domain expertise with machine learning will need refinement. For example, algorithms often rely on three types of anomalies to detect fraud: those based on outliers in a distribution (e.g., transactions above a certain threshold), those that are anomalous because of the context (e.g., a high withdrawal in a period of low economic activity), and those that are anomalous when multiple observations are analyzed jointly (e.g., ATM transactions from two continents over a short period of time). The first type of anomalies can probably be detected with minimal domain expertise. However, the second and third types often require domain expertise to be effective.

Further, analytical processes must be analyzed and maintained regularly, adding layers of complexity in the organizational system. Other types of failures in the control system (e.g., cyber-risk) will have to be considered. Privacy issues (with different legal standards across jurisdictions) are likely to become increasingly important. All the requirements necessitate new skills that may be in short supply at the moment. This may explain why, according to some statistics, 85% of AI projects fail.[22]

---

[22] https://customerthink.com/why-85-of-the-artificial-intelligence-projects-fail/

Despite all the challenges, we believe that using machine learning to combat fraud remains a promising direction of inquiry for both researchers and industry practitioners. We look forward to more innovative approaches from future research to tackle this important problem facing our societies.

REFERENCES

Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). Metafraud: a meta-learning framework for detecting financial fraud. *Mis Quarterly*, 1293-1327.

American Institute of Certified Public Accountants (2002) Consideration of Fraud in a Financial Statement Audit. Statement on Auditing Standards No. 99. New York.

Amiram, D., Bozanic, Z., & Rouen, E. 2015. Financial statement errors: Evidence from the distributional properties of financial statement numbers. *Review of Accounting Studies*, 20, 1540–1593.

Ashton, R. H. (1974). Behavioral implications of information overload in managerial accounting reports. *Cost and Management*, 48(4), 37-40.

Baltrušaitis, T., Ahuja, C., & Morency, L.P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2), 423-443.

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting Accounting Fraud in Publicly Traded US Firms Using a Machine Learning Approach. *Journal of Accounting Research*, 58(1), 199-235.

Beasley, M. S. 1996. An empirical analysis of the relation between the board of director composition and financial statement fraud. *The Accounting Review*, 71, 443–465.

Beasley, M. S., J. V. Carcello, and D. R. Hermanson. 1999, Fraudulent Financial Reporting: 1987-1997: An Analysis of U.S. Public Companies. Sponsored by the Committee of Sponsoring Organizations of the Treadway Commission (COSO).

Beasley, M. S., J. V. Carcello, D. R. Hermanson, and T. L. Neal. 2010. Fraudulent Financial Reporting: 1998-2007: An Analysis of U.S. Public Companies." Sponsored by the Committee of Sponsoring Organizations of the Treadway Commission (COSO).

Bekker, J., & Davis, J. (2020). Learning from positive and unlabeled data: a survey. *Mach. Learn.*, 109(4), 719-760.

Beneish, M. D. "Detecting GAAP Violation: Implications for Assessing Earnings Management among Firms with Extreme Financial Performance." *Journal of Accounting and Public Policy* (16) (1997): 271-309.

Beneish, M. D. "The Detection of Earnings Manipulation." *Financial Analysts Journal* (55) (1999): 24-36.

Benbasat, I., & Taylor, R. N. (1982). Behavioral aspects of information processing for the design of management information systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 12(4), 439-450.

Beutel, A., Akoglu, L., & Faloutsos, C. (2015) 'Graph-based user behavior modeling: from prediction to fraud detection' *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. pp. 2309-2310.

Brazdil, P., Carrier, C. G., Soares, C., & Vilalta, R. (2008). Metalearning: Applications to data mining. Springer Science & Business Media.

Brazel, J. F., K. L. Jones, and M. F. Zimbelman. "Using nonfinancial measures to assess fraud risk." *Journal of Accounting Research* 47 (5) (2009): 1135–66.

Brown, N.C., Crowley, R.M. and Elliott, W.B. 2020, What Are You Saying? Using topic to Detect Financial Misreporting. *Journal of Accounting Research*, 58:

237-291.

Burns, N., & Kedia, S. 2006. The impact of performance-based compensation on misreporting. *Journal of Financial Economics*, 79, 35–67.

Cao, S., Yang, X., Chen, C., Zhou, J., Li, X., & Qi, Y. (2019). TitAnt: online real-time transaction fraud detection in Ant Financial. *arXiv preprint arXiv:1906.07407*.

Chen, X., Hilary, G. and Tian, X. (2020), Mandatory Data Breach Transparency and Insider Trading, working paper.

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Making words work: Using financial text as a predictor of financial events. *Decision Support Systems*, 50(1), 164-175.

Citron, D. K. (2008). Technological due process. *Wash. UL Rev.*, 85, 1249.

Darrough, M., Huang, R. and Zhao, S. (2020), Spillover Effects of Fraud Allegations and Investor Sentiment. *Contemporary Accounting Research*, 37: 982-1014.

Davidson, R., Dey, A., & Smith, A. 2015. Executives' Boff-the-job^ behavior, corporate culture, and financial reporting risk. *Journal of Financial Economics*, 117(1), 5–28.

de Roux, D., Perez, B., Moreno, A., Villamil, M. d. P., & Figueroa, C. (2018) 'Tax fraud detection for under-reporting declarations using an unsupervised machine learning approach' *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. pp. 215-222.

Dechow, P. M., R. G. Sloan, and A. P. Sweeney. "Detecting earnings management." The Accounting Review 70 (2) (1995): 193–226.

Dechow, P. M., R. G. Sloan, and A. P. Sweeney. "Causes and Consequences of Earnings Manipulation: An Analysis of Firms Subject to Enforcement Actions by the SEC." Contemporary Accounting Research 13 (1996): 1-36.

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting material accounting misstatements. Contemporary accounting research, 28(1), 17-82.

Dong, W., Liao, S., & Zhang, Z. (2018). Leveraging financial social media data for corporate fraud detection. *Journal of Management Information Systems*, 35(2), 461-487.

Dutta, I., Dutta, S., & Raahemi, B. (2017). Detecting financial restatements using data mining techniques. *Expert Systems with Applications*, 90, 374-393.

Dyck, A., A. Morse and L. Zingales. 2020. How pervasive is corporate fraud. University of Toronto working paper.

Efendi, J., A. Srivastava, and E. P. Swanson. "Why do corporate managers misstate financial statements? The role of option compensation and other factors" Journal of Financial Economics (85) (2007): 667-708.

Ernst & Young. 2010. Driving ethical growth—New markets, new challenges. 11th Global Fraud Survey. Available online at https://linomartins.files.wordpress.com/2011/12/2011th_global_fraud_survey.pdf.

Fawcett, T. "An Introduction to Roc Analysis." *Pattern Recognition Letters* (27) (2006): 861-874.

Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, 479, 448-455.

Fletcher H., Glancy, Surya B. Yadav, 2011. A computational model for financial reporting fraud detection, *Decision Support Systems*, 50 (3), 595-601.

Garip, F. (2020). What failure to predict life outcomes can teach us. *Proceedings of the National Academy of Sciences*, 117(15), 8234-8235.

Green, P., and J. H. Choi. "Assessing the Risk of Management Fraud through Neural Network Technology." *Auditing: A Journal of Practice & Theory* (16) (1997): 14–29.

Guo, J., Liu, G., Zuo, Y., & Wu, J. (2018) 'Learning sequential behavior representations for fraud detection' *2018 IEEE international conference on data mining (ICDM)*. IEEE, pp. 127-136.

Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud–A comparative study of machine learning methods. *Knowledge-Based Systems*, 128, 139-152.

Hastie, T., R. Tibshirani, and J.H. Friedman. "The Elements of Statistical Learning." New York: Springer, 2009.

He, H., and Y. Ma. "Imbalanced Learning: Foundations, Algorithms, and Applications." *Wiley*, 2013.

Healy, P. M. (1985). The effect of bonus schemes on accounting decisions. Journal of Accounting and Economics, 7(1), 85–107.

Hobson, J. L., Mayew, W. J., & Venkatachalam, M. (2012). Analyzing speech to detect financial misreporting. *Journal of Accounting Research*, 50(2), 349–392.

Hoi, S. C., Sahoo, D., Lu, J., & Zhao, P. (2018). Online learning: A comprehensive survey. *arXiv preprint arXiv:1802.02871*.

Hu, B., Zhang, Z., Shi, C., Zhou, J., Li, X., & Qi, Y. 33 (2019) 'Cash-out user detection based on attributed heterogeneous information network with a hierarchical attention mechanism' *Proceedings of the AAAI Conference on Artificial Intelligence*. pp. 946-953.

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, 50(3), 585-594.

Iselin, E. R. (1988). The effects of information load and information diversity on decision quality in a structured decision task. *Accounting, organizations and Society*, 13(2), 147-164.

Järvelin K. and J. Kekäläinen. "Cumulated Gain-Based Evaluation of IR Techniques." *ACM Transactions on Information Systems* (20) (2002): 422-446.

Johnson, S. A., H. E. Ryan, and Y. S. Tian. "Managerial Incentives and Corporate Fraud: The Sources of Incentives Matter." Review of Finance (13) (2009): 115-145.

Karpoff, J. M., D. S. Lee, and G. S. Martin. 2008. The costs to firms of cooking the books. *Journal of Financial and Quantitative Analysis* 43 (03): 581–612.

Karpoff, J.M., Koester, A., Lee, D.S., & Martin, G.S. 2017. Proxies and databases in financial misconduct research. *The Accounting Review*, 92(6), 129–163.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer. "Prediction Policy Problems." *American Economic Review: Papers & Proceedings*, 105(5), (2015): 491–495.

KPMG Peat Marwick. 1998. Fraud Survey, KPMG Peat Marwick, Montvale,

NJ.

Larcker, D. F., Richardson, S. A., & Tuna, I. 2007. Corporate governance, accounting outcomes, and organizational performance. *The Accounting Review*, 82(4), 963–1008.

Larcker, D. and A. A. Zakolyukina. "Detecting Deceptive Discussion in Conference Calls." *Journal of Accounting Research* (50) (2012): 495-540.

Li, H., Liu, B., Mukherjee, A., & Shao, J. (2014). Spotting fake reviews using positive-unlabeled learning. *Computación y Sistemas*, 18(3), 467-475.

Liang, C., Liu, Z., Liu, B., Zhou, J., Li, X., and Yang, S.,(2019) 'Uncovering Insurance Fraud Conspiracy with Network Learning' *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 1181-1184.

Lin, Jerry & Hwang, Mark & Becker, Jack. (2003). A fuzzy neural network for assessing the risk of fraudulent financial reporting. *Managerial Auditing Journal*. 18. 657-665.

Liu, S., Hooi, B., & Faloutsos, C. (2019). A contrast metric for fraud detection in rich graphs. *IEEE Transactions on Knowledge and Data Engineering*, 31(12), 2235-2248.

Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, 50(3), 559-569.

Oentaryo, R., Lim, E.-P., Finegold, M., Lo, D., Zhu, F., Phua, C., et al. (2014). Detecting click fraud in online advertising: a data mining approach. *The Journal of Machine Learning Research*, 15(1), 99-140.

Perols, J. L., R. M. Bowen, C. Zimmermann, and B. Samba. (2017). Finding Needles in a Haystack: Using Data Analytics to Improve Fraud Prediction. The Accounting Review (92), 221-245.

Purda, L., & Skillicorn, D. (2015). Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research*, 32(3), 1193-1223.

Salganik, M., Lundberg, I., Kindel, A., Ahearn, C., Al-Ghoneim, K. Almaatouq, A., Altschul, D., Brand, J., Carnegie, N., Compton, R, Datta, D., Davidson, T., Filippova, A., Gilroy, C., Goode, B., Jahani, E., Kashyap, R., Kirchner, A., Mckay, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. Proceedings of the National Academy of Sciences. 117.

Shah, N., Lamba, H., Beutel, A., & Faloutsos, C. (2017) 'The many faces of link fraud' *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, pp. 1069-1074.

Shmueli, G. "To explain or to predict." *Statistical Science* (25) (2010): 289-310.

Van Vlasselaer, V., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2017). Gotcha! Network-based fraud detection for social security fraud. *Management Science*, 63(9), 3090-3110.

Varian, H.R. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* (28) (2014): 3–28.

Wang, D., Lin, J., Cui, P., Jia, Q., Wang, Z., Fang, Y., et al. (2019) 'A Semi-supervised Graph Attentive Network for Financial Fraud Detection' *2019 IEEE*

International Conference on Data Mining (ICDM). IEEE, pp. 598-607.

Wang Y., Wang L., Li Y., He D., Chen W., Liu T.-Y. "A Theoretical Analysis of NDCG Ranking Measures." In *Proceedings of the 26th Annual Conference on Learning Theory*, 2013.

Wang, J., Wen, R., Wu, C., Huang, Y., & Xion, J. (2019) 'Fdgars: Fraudster detection via graph convolutional networks in online app review system' *Companion Proceedings of The 2019 World Wide Web Conference*. pp. 310-316.

Wang, Y., & Xu, W. (2018). Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems*, 105, 87-95.

Whiting, D.G., Hansen, J.V., McDonald, J.B., Albrecht, C. and Albrecht, W.S. (2012), machine learning methods for detecting patterns of management fraud. *Computational Intelligence*, 28: 505-527.

Xu, C., Zhang, J., & Sun, Z. (2017) 'Online Reputation Fraud Campaign Detection in User Ratings' *IJCAI*. pp. 3873-3879.

Yuan, S., Wu, X., Li, J., & Lu, A. (2017) 'Spectrum-based deep neural networks for fraud detection' *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. pp. 2419-2422.

Zhang, J., Yang, X., & Appelbaum, D. (2015). Toward effective Big Data analysis in continuous auditing. *Accounting Horizons*, 29(2), 469-476.

Zhang, Y.-L., Zhou, J., Zheng, W., Feng, J., Li, L., Liu, Z., et al. (2019). Distributed deep forest and its application to automatic detection of cash-out fraud. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(5), 1-19.

Zheng, P., Yuan, S., Wu, X., Li, J., & Lu, A. 33 (2019) 'One-class adversarial nets for fraud detection' *Proceedings of the AAAI Conference on Artificial Intelligence*. pp. 1286-1293.

Zhong, Q., Liu, Y., Ao, X., Hu, B., Feng, J., Tang, J., et al. (2020) 'Financial Defaulter Detection on Online Credit Payment via Multi-view Attributed Heterogeneous Information Network' *Proceedings of The Web Conference 2020*. pp. 785-795.

Zhu, Y., Xi, D., Song, B., Zhuang, F., Chen, S., Gu, X., et al. (2020) 'Modeling Users' Behavior Sequences with Hierarchical Explainable Network for Cross-domain Fraud Detection' *Proceedings of The Web Conference 2020*. pp. 928-938.