Diginomics
WORKING
PAPER

Research Group
Diginomics
Digitalization - Economy - Society

Machine Learning in Accounting Research

Christian Fieberg, Matthies Hesse,
Thomas Loy, Daniel Metko

April 2022

No. 0014

Universität Bremen

Impressum:

Diginomics Working Paper

ISSN: 2701-6307

DOI: http://dx.doi.org/10.26092/elib/1511

Published by:
Universität Bremen, Diginomics Research Group, Max-von-Laue-Straße 1, 28359 Bremen, Germany

Editor:
Prof. Dr. Lars Hornuf
Phone: +49(0)421 218 66820
E-mail: hornuf@uni-bremen.de

https://www.uni-bremen.de/graduiertengruppe-diginomics/forschung/diginomics-working-paper

This working paper will appear in L. Hornuf (Ed.), *Diginomics Research Perspectives: The Role of Digitalization in Business and Society*. Cham: Springer International Publishing.

Machine Learning in Accounting Research

Christian Fieberg, Matthies Hesse, Thomas Loy and Daniel Metko

Abstract

We present a compact overview of machine learning applications in financial accounting and audit research as well as management accounting research. Here, the application of machine learning has the potential to provide novel insights into empirical data and to improve predictive performance. We highlight the potential use of deep learning to process unstructured and structured data more efficiently and a greater focus on model interpretability as viable opportunities for future research.

6.1 Introduction

Artificial intelligence (AI), as a generic term for machine learning (ML), is already an indispensable part of our daily personal and business lives. Consumers enjoy personalized film and music recommendations on various platforms as well as the help of their voice-activated digital assistants. Businesswise, manufacturers trust in automatic quality control systems to detect defective parts, banks predict credit default probabilities to mitigate risks, doctors retrieve second opinions from artificial intelligence systems in critical on-demand situations and retailers optimize their stocks on predicted demand (Kashyap, 2017). All of this is thanks to ML. The widespread use of ML seems ever-increasing and does not stop at accounting and empirical accounting research. Empirical accounting research consists of two main areas: (1) financial accounting and auditing, and (2) management accounting. Financial accounting, also referred to as external accounting or financial reporting, is the communication of complex, aggregated business transactions to a wide set of internal and external stakeholders, such as investors, creditors, suppliers, customers and employees. Depending on the jurisdiction, taxable income might also be deferred from the profit and loss statement contained in an overall (annual) financial statement. In contrast, auditing is the process of validating these financial statements such that they comply with the respective generally accepted accounting principles (GAAP). Management accounting, also referred to as internal accounting, has a greater focus on value-creation by, for example, estimating costs, setting budgets, or evaluating future project profitability. It is especially useful for strategic decision-makers inside of firms.

We identify the current state of use of ML along these main lines of accounting research. Specifically, we find that in financial accounting and auditing, ML is used in such areas as fraud detection and bankruptcy prediction. Additionally, we find that ML remains a rather niche topic in management accounting. We contribute to the literature by providing a compact overview of ML and its use in various areas of accounting research. We also identify research opportunities, such as: (1) more extensive use of deep learning to process structured and unstructured data more efficiently, and (2) use of ML as a white-box application (i.e., explainable ML) in contrast to the proverbial black-box of ML. In the first part of this chapter, we provide a brief introduction to ML by clarifying its meaning, showing popular techniques and presenting exemplary use cases. With this knowledge, the reader can easily follow the subsequent discussion, where we outline the use of ML in empirical accounting research. This allows us to appraise the current state of ML and to identify potential avenues for future research.

6.2 Machine Learning

6.2.1 Cutting Through the Fog

The term artificial intelligence (AI) has long since become a buzzword. In the popular press, it often goes hand-in-hand with such terms as business analytics, big data, business intelligence, data analytics, data mining, statistical learning, machine learning (ML), and deep learning. Although these terms are used very frequently and seemingly interchangeably, we aim to disentangle them before taking a deeper look at the techniques behind ML and its application in the narrower context of accounting research.

Business analytics refers to the analysis of data (i.e., data analytics) within a business. Business data mostly meet the criteria of the term big data, with its three main characteristics: *volume*, *velocity*, and *variety*. Volume refers to the amount of data, velocity to the speed of the data generation and processing, and variety to the number of different data types (i.e., structured, unstructured and semi-structured data) (Lee, 2017). Structured data is the kind that can be found in traditional, relational databases, whereas unstructured data are not standardized for efficient processing. Examples of unstructured data are texts, photos, audio and audio-visual data. Semi-structured data are not standardized in the pure sense of a relational database, but still structured. An example would be financial statements in which unstructured data (e.g., texts contained in a footnote disclosure or in the Management Discussion & Analysis (MD&A) section) are electronically tagged using the eXtensible Business Reporting Language (XBRL), making them retrievable automatically (e.g., Hoitash & Hoitash, 2018). Business analytics is then used to analyze the business-generated data to support decision-making.

While it has a rather descriptive character, business intelligence focuses primarily on prediction tasks. Business intelligence uses data analytics – or, more specifically, data mining – to detect patterns and gain knowledge from the data (Mishra et al., 2016). Data mining uses statistical learning and machine learning, both of which are predictive toolboxes. Statistical learning is an emergent field from statistics, while machine learning has a long history within computer science (James et al., 2013). Whereas statistical learning makes predictions based on statistical assumptions, machine learning acts in a purely data-driven way – the "machine" learns relationships within the data itself (i.e., self-learning) and predicts an outcome. Finally, deep learning is a subcategory of ML in which a neural network with many hidden layers at its core allows for deeper insights into the input–output relation structure. In other words, deep learning is more flexible in detecting hidden patterns than "classical" ML techniques such as decision trees, support vector machines or neural networks with only one hidden layer. Throughout this chapter, we will subsume both statistical learning and machine learning under the term ML for readability purposes.

6.2.2 Supervised and Unsupervised Learning Techniques

ML can generally be categorized into supervised, unsupervised, semi-supervised, and reinforcement learning (e.g., Géron, 2017). For the sake of brevity, this chapter focuses on supervised and unsupervised learning, but the concept of semi-supervised learning is introduced along the way.

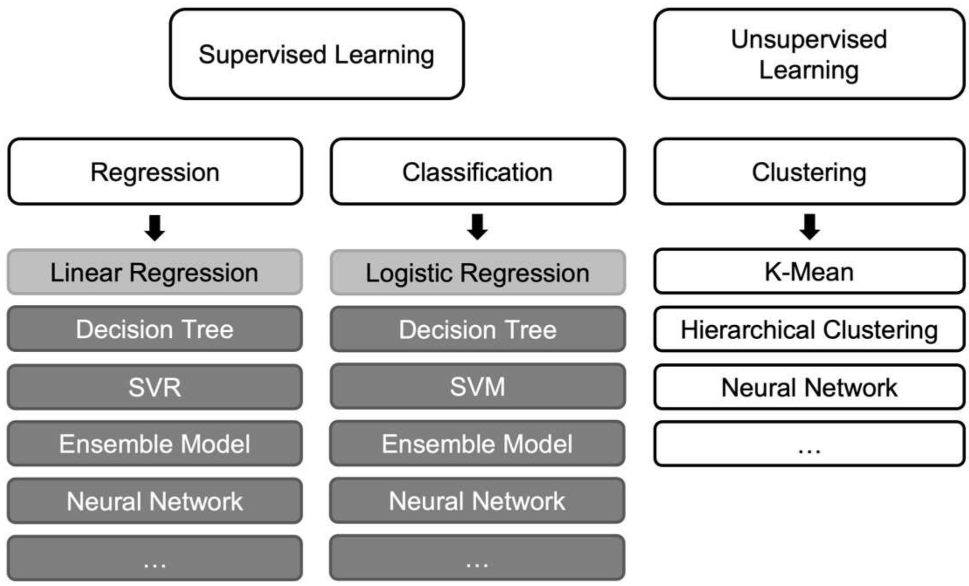

Fig 0.1 *Exemplary algorithms of supervised and unsupervised learning*

With supervised learning, the main idea is to discover and approximate the relation between given input and output variables. In statistics, these input and output variables are called independent and dependent variables. Within ML, they are referred to as features and targets.[1] One difference between (classical) statistics and ML is that statisticians aim to derive inferences while ML engineers aim to predict the targets as accurately as possible. As this chapter is about ML in accounting research and, therefore, largely focuses on prediction tasks, ML notation is employed throughout this chapter. In terms of data, supervised learning is only applicable if each observation contains a number of features as well as corresponding target(s). Supervised learning is sub-divided into regression and classification, depending on the targets' data type. A regression algorithm is used if targets are continuous, while a classification algorithm requires categorical targets.

Fig 0.1 shows some popular algorithms. Algorithms marked in light gray are only able to learn linear approximation functions (in the context of regression) or decision boundaries (in the context of classification). Examples of approximation functions and decision boundaries are provided in the following section. The algorithms presented in dark gray can be adapted such that they are able to learn linear or non-linear relations within the data. Support vector regression (SVR), support vector machines (SVMs), and neural networks are, for example, applicable in both linear and non-linear prediction tasks, with slight adjustments. Decision trees and more complex variations such as random forests and gradient-boosted trees, combined within ensemble models (i.e., combinations of predictions of individually poor performing models – so called "weak learners"), are inefficient in learning linear relations. The advantage of linear algorithms is that the results are easier to interpret. On the other hand, algorithms with non-linear learning abilities are more flexible and can therefore detect non-linearities with ease, but are also prone to overfit. Overfitting means that a model performs well on data used in the training process (i.e., training data) but poorly on novel (i.e., test) data.

[1] Sometimes, targets are synonymously referred to as "labels", "annotations", "responses" or simply "outputs".

Most real-world data is available in unstructured form and a corresponding target is mostly not assigned (Lee, 2017). Without these feature–target pairs, supervised learning techniques are infeasible. However, unsupervised learning can extract meaningful information from unstructured data. **Fig 0.1** summarizes some popular unsupervised learning algorithms. In general, unsupervised learning techniques cluster unstructured data to either draw conclusions or to assign targets which can then be used for supervised learning. Lastly, semi-supervised learning combines supervised and unsupervised learning by having a small set of feature–target pairs and a big set of features not assigned to targets (for example, to boost predictive performance). Here, a model is trained on the small set of available feature–target pairs in the first step. This model is then used to predict targets for the big set of features initially not assigned to targets. The model can then be retrained with all feature–target pairs (for an application in credit card fraud detection, see Melo-Acosta et al. 2017).

6.2.3 The Machine Learning Research Process

ML research consists of six distinct phases: (1) data extraction, (2) data preprocessing, (3) data analysis, (4) data mining, (5) result evaluation and (6) result interpretation. Although data mining (i.e., the actual use of ML algorithms) is only a fraction of the entire research process, finding the most suitable ML algorithm and applying it properly can be tedious and time-consuming. Specifically, the selection of an ML algorithm is not trivial. There is no generally superior ML algorithm, as algorithmic performance depends solely on the given data (Goodfellow et al., 2016). Because supervised learning has been most commonly applied within accounting research, the discussion below is confined to supervised learning. Unsupervised learning is used rather sporadically or as a first step in a two-step research approach combined with supervised learning (e.g., to assign unstructured data to targets which are then used as features for a supervised learning task; Brown et al., 2020). Every supervised learning technique approximates a function $f(.)$ that maps the features X onto the corresponding target y:

$$y \approx f(X).$$

In a regression task, y takes on a continuous value. In a binary classification task, y is restricted to two values (e.g., zero and one). In the latter, it is convenient to write y as a probability of X, $p(X)$. To visualize the mapping process of ML techniques with linear and non-linear learning abilities, Figures 6.2 and 6.3 present examples of regression and classification, respectively.

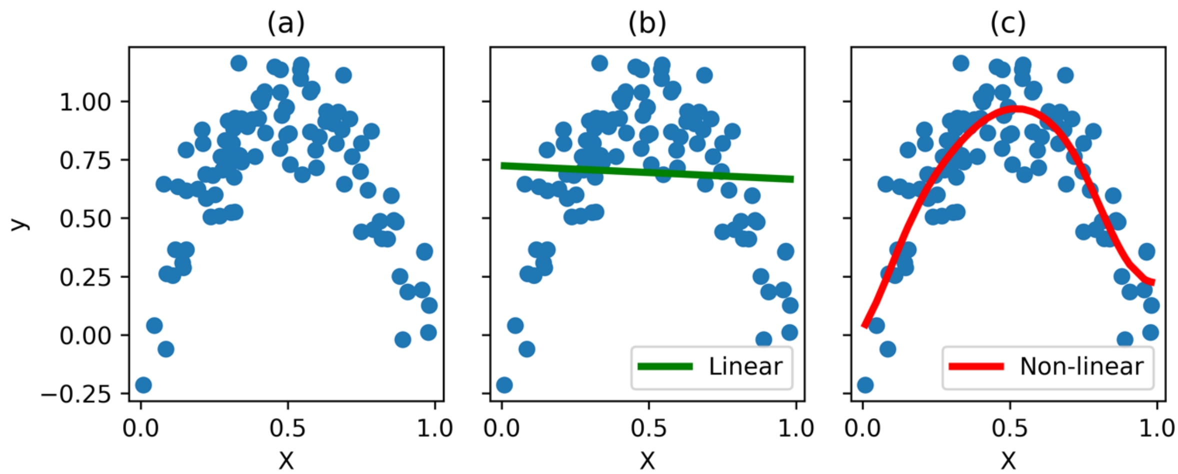

Fig. 0.2 *Regression example with 100 observations, feature values on the x-axis and the corresponding targets on the y-axis*

Fig. 0.2 (a) illustrates a regression example with 100 observations, with feature values presented on the x-axis and target values on the y-axis. As one can see, the dots follow an inverted parabola (i.e., a non-linear pattern). Therefore, it is no surprise that an ML algorithm that is only able to learn linear relations between features and targets does not approximate the patterns in the data very well – see **Fig. 0.2** (b) where a linear regression is used for approximation. An ML algorithm that is able to learn non-linearities within the data, such as SVR, is thus eminently superior in this case, as depicted in **Fig. 0.2** (c).

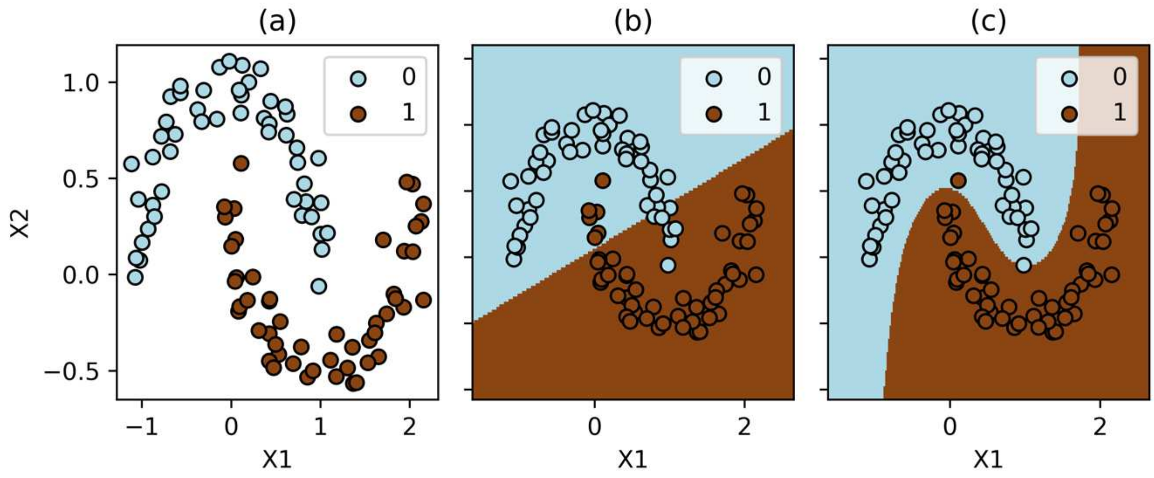

Fig. 0.3 Binary classification example with 100 observations, feature values on the x- and y-axis and colored targets

In the case of a classification task, the algorithm needs to find a decision boundary which separates the classes from each other. **Fig. 0.3** (a) represents a binary classification example, again with 100 observations. In this example, there are two features, one on the x-axis and one on the y-axis. The corresponding class is visualized by the color of the dots. Similar to the regression task, one can see that a linear class separation is not optimal. **Fig. 0.3** (b), in which a logistic regression is used, confirms this. The background color corresponds to the class prediction of each dot in the graph. However, an ML algorithm with non-linear learning abilities, such as the SVM, separates the two classes almost perfectly (cf., **Fig. 0.3** (c)).

6.3 Machine Learning in Accounting Research

As discussed above, ML techniques with non-linear learning abilities extend ML techniques with linear learning abilities by increasing the models' ability to detect non-linearities and interdependent patterns that may uncover otherwise hidden dependencies between given feature and target variables. Therefore, ML opens new research opportunities which also benefit accounting research. We structure our review of the literature along the main lines of ML applications in accounting research, being (1) financial accounting and auditing, and (2) management accounting.

6.3.1 Financial Accounting and Auditing

Financial accounting refers to the preparation of financial statements, whereas auditing is concerned with the verification of (annual) financial statements. The objective of an audit is to confirm with reasonable assurance that a financial statement complies with generally accepted accounting principles (GAAP). As this section demonstrates, comparatively more work has been done with respect to ML in audit research-related topics than in financial accounting.

As an example, Baldwin and colleagues (2006) discuss some major audit tasks suitable for AI based on an earlier classification of these tasks by Abdolmohammadi (1991). They show that many of the 332 audit tasks (e.g., computation of various inventory ratios; comparative studies of sales, bad debts, cash receipts, warranties, and returns) which Abdolmohammadi (1991) deemed suitable for employing decision aid systems are also amenable to automation through ML. More specifically, Baldwin and colleagues (2006) identify classification tasks like differentiating between collectible and bad debts or legitimate and questionable transactions, and risk-assessment tasks like misstatement detection or going-concern predictions. Following Baldwin and colleagues (2006), there have been renewed calls for more research on the possibilities of AI in auditing (e.g., Brandas et al., 2018; Cao et al., 2015; Earley, 2015; Gepp et al., 2018). Besides the studies already mentioned, many more provide an overview on the application of AI in the field of auditing (e.g., IAASB, 2016; Kokina & Davenport, 2017; T. Sun, 2019; T. Sun & Vasarhelyi, 2017). The main takeaway is that ML makes auditing more efficient and effective because it enables auditors to deal with large amounts of structured and unstructured data which would otherwise be impossible to process.

The literature often suggests ML for specific audit tasks like going-concern prediction or misstatement detection instead of regarding the entire audit process as a potential ML application. As an example for the latter, No and colleagues (2019) present a potential integration of ML into a larger part of the entire audit process in the form of an ML-based sampling technique. Auditors commonly resort to sampling a (small) percentage of the population of auditable items (e.g., assets, transactions). As a second step, audit findings from the sample items are extrapolated onto the full population. One can imagine that this procedure is not always fail-safe for achieving reasonable assurance for the overall financial statement. Therefore, the authors present a framework called Multidimensional Audit Data Selection (MADS) which is based on ML. First, an auditor trains the AI by filtering items that have higher risks of being misstated out of the entire population. Then, based on these items, an ML outlier detection technique is used to identify further outliers. Afterwards, the outliers are prioritized by a weighting scheme to narrow down the auditor's examination set. MADS improves the efficiency and effectiveness of an auditor by highlighting the transactions with a higher risk of material misstatement (No et al., 2019). Such an approach remains rather rare in the audit research literature; in the following sections, we turn to ML-supported audit tasks that have gained much more traction recently – namely, the detection of fraud and the prediction of going-concern or bankruptcy.

6.3.2 Fraud Detection

Accounting fraud is a global problem, as seen in the many high-profile cases such as Enron in the US, Sino-Forest in China, Vivendi Universal in France and, more recently, Wirecard in Germany. There are estimates that fraud causes global financial losses of US$4.5 trillion each year (Association of Certified Fraud Examiners, 2020). Auditors are responsible for ensuring that financial statements are free from material misstatements originating from either (unintentional) error or (intentional) fraud. However, Fanning and Cogger (1998) state that most auditors possess a lack of knowledge when it comes to the characteristics of accounting fraud and that they are not experienced enough to detect it. Additionally, they point out that managers involved in accounting fraud try to deceive auditors. Besides being the result of intentional misstatements, fraud can be further divided into fraudulent financial reporting (i.e., management fraud) and misappropriation of assets (i.e., employee theft/fraud). Examples of fraudulent financial reporting would be the "manipulation, falsification or alteration of accounting records or supporting documents from which financial statements are prepared or misrepresentation in or intentional omission from the financial statements of events, transactions, or other significant information" (SAS No. 99 Par. 6). Misappropriation of assets involves "the theft of an entity's assets where the effect of the theft causes the financial statements not to be presented, in all material respects, in conformity with GAAP" (SAS No. 99 Par. 6). There is evidence that misappropriation of assets occurs eight times more often than financial statement fraud, but financial statement fraud results in far greater financial losses (Association of Certified Fraud Examiners, 2020). This could be one reason for the greater academic attention to financial statement fraud detection, as discussed below. A commonly used target dataset for financial statement fraud detection in the US is the SEC's Accounting and Auditing Enforcement Releases (AAERs).[2]

The use of ML to detect financial statement fraud can be justified by the associated improvements in auditor decisions and regulators' abilities to identify potentially fraudulent firms by having (more accurate) fraud detection systems (e.g., Hooda et al., 2018; SEC, 2015). By construction, fraud detection is a supervised classification task. Before the application of ML, most research in this area thus relied on logistic regression (e.g., Dechow et al., 2011), which still acts as a popular benchmark in contemporaneous studies (e.g., Bao et al., 2020a). The related literature relies heavily on publicly available financial data in structured form (i.e., financial statement numbers; e.g., Cecchini et al., 2010a; Dechow et al., 2011; Green & Choi, 1997). Nevertheless, unstructured data are used as well, mostly in textual form extracted from the MD&A sections of the financial reports (e.g., Cecchini et al., 2010b; Humpherys et al., 2011; Purda & Skillicorn, 2015). Other unstructured data source examples are social media for textual data or earnings conference calls for auditive data (Dong et al., 2018; Hobson et al., 2012). Methods of processing textual data include counting words that are predictive of fraudulent behavior or analyzing sentiments contained in social media data.[3] This information is then used as features for the classification task. Along with the general challenges of ML (e.g., finding general patterns within the data without overfitting), fraud detection comes with its own challenges. Fraud, in general, is rare and conducted adversarially, meaning fraudsters try their hardest to prevent learning from past events (Bao et al., 2020a). The (relative) scarcity of fraud compared to non-fraud cases results in an uneven ratio between fraud and non-fraud cases in any dataset. This issue is also known as the "imbalance problem." Fraud is not only rare, but it is often not detected, and if so, only years later (Bao et al., 2020a). Additionally, fraud is rarely a single event but often preceded by a time-series of violations or a slippery slope beginning with minor infractions, despite most research handling fraud as single one-off events. Lastly, misclassification of fraudulent firms is more costly than misclassification of non-fraudulent firms. In line with the respective literature (Bao et al., 2020a; Perols, 2011), we refer to this as the "cost problem."

Initial studies linking fraud detection and ML employ neural networks (e.g., Fanning & Cogger, 1998; Green & Choi, 1997). For example, Green and Choi (1997) construct a sample with an even number of fraud and non-fraud cases extracted from the AAER dataset. They report an improvement of fifty percent by using the neural network

[2] AAERs contain financial statements with fraudulent account balances identified within two years after the audited financial statement was released (e.g., Green & Choi, 1997).

[3] For more information about social media analytics, see Kinra and colleagues (2022).

instead of random choice.[4] Fanning and Cogger (1998) use a much bigger target and feature set. They consider twenty potential predictors, including some related to the firms' governance structure, such as board size, percentage of outside directors and the duality of roles between chairperson of the board and CEO. With the increased popularity of ML, many more classification methods are used and compared. Kotsiantis and colleagues (2006) use financial ratios to predict fraud within manufacturing firms listed on the Athens Stock Exchange. Instead of just using individual classification methods and comparing their results, they build a classification ensemble. More specifically, they train individual classification models and combine their predictions according to a voting scheme.[5] Among others, they use classification models based on decision trees, neural networks, logistic regressions, and SVMs. With this approach, they outperform the prediction of each individual ML algorithm.

Perols (2011) critizes the prior literature for not appropriately considering the challenges of fraud detection. He uses undersampling[6] to solve the imbalance and cost problem. In an experimental setting, the author determines the best undersampling ratio (i.e., ratio between fraud and non-fraud cases) in the training and test dataset. Perols (2011) uses similar algorithms as Kotsiantis and colleagues (2006). A subsequent study by Perols and colleagues (2017) employs two more methods to adress the imbalance problem: (1) the abdundance of already identified features in the literature, which can result in the "curse of dimensionality,"[7] and (2) the common handling of fraud as homogeneous events, although one could differentiate between, for example, revenue and expense fraud. Besides undersampling the observations, they use two variations of undersampling on the feature level: randomly selecting features and selecting features based on *a priori* knowledge to group them into specific types of fraud, e.g., revenue vs. expense fraud (Perols et al., 2017). Observation undersampling and feature undersampling with *a priori* knowledge work best under the out-of-sample (OOS) approach[8] and result in an accuracy improvement of about ten percent compared to the benchmarks (Perols et al., 2017). Rather than changing the sample composition, Kondo and colleagues (2019) use an ensemble model that still performs well on imbalanced data. They incorporate a weighted random forest which penalizes misclassifications of fraudulent cases more than of non-fraudulent cases (Chen et al., 2004).

Li and colleagues (2016) are the first who used raw accounting variables (i.e., publicly available financial data) instead of financial ratios (e.g., modified Jones model discretionary accruals) or non-financial ratios (e.g., abnormal change in employees, Dechow et al., 2011). While raw accounting variables form the basis for financial

---

[4] Of course, random choice would result in a fifty-percent chance of making the correct prediction due to the balanced dataset.

[5] For example, counting the classification predictions of each model and selecting the majority class.

[6] Undersampling means randomly deleting cases from the majority class (i.e., non-fraudulent cases in the context of fraud detection). In contrast, oversampling results in the duplication of cases in the minority class (i.e., fraudulent cases in the context of fraud detection). Undersampling, as compared to oversampling, is more effective against class and cost imbalances (Drummond & Holte, 2003).

[7] The curse of dimensionality describes that a complete search for an optimal solution on the entire feature space is becoming computationally harder with every additional feature (Bellman, 1961).

[8] Target prediction on novel data, i.e., data not used in the training process.

and non-financial ratios, they argue that raw accounting variables contain more information because ratios lose information when they get constructed. They use SVMs with different kernel functions for the prediction and ensemble methods. It is worth noting that they specifically use an OOS procedure instead of an in-sample (IS) approach that is commonly used for causal inferences (Li et al., 2016). IS results in very optimistic predictions and is not recommended because predictions are made on data already seen by the model; hence, the more the model overfits, the better the predictions. They show superior performance for all methods using raw accounting variables to the benchmarks with financial and non-financial ratios (Li et al., 2016). Bao and colleagues (2020b) conduct a similar study, but they compare more ML methods and use different evaluation metrices. The predictive power of unstructured, textual data from the MD&A section in combination with financial ratios is studied by Craja and colleagues (2020). They present a deep learning model to benefit from unstructured and structured data at the same time. The authors report a substantial predictive improvement compared to commonly used ML techniques such as random forests or SVMs. Furthermore, the deep learning architecture allows them to look into the "black-box" and to identify early warning indicators on the word- and sentence-level. On the word-level, they report early warning indicators like "cost" or "acquisition" and on the sentence-level sentences like "If we are unable to fully integrate acquired products, technologies or businesses, or train, retain and motivate personnel from the acquired businesses, we may not receive the intended benefits of those acquisitions, which could seriously harm our business, operating results and fiancial condition."

Many studies try to detect fraudulent financial statements, but only a few also take errors (i.e., unintentional misstatements) under consideration. Dutta and colleagues (2017) fill this research gap by forecasting erroneous as well as fraudulent financial statements. They incorporate a two-step approach consisting of feature selection (i.e., removal of less significant and redundant features) and restatement prediction based on algorithms like decision trees, neural networks, and SVMs. Bertomeu and colleagues (2020) extend this research by replacing the two-step approach with a regression tree ensemble model, referred to as gradient-boosted regression trees, and the incorporation of additional features.

6.3.3 Bankruptcy Prediction

Bankruptcy prediction, sometimes also referred to as financial distress prediction, has a long history in accounting and finance research, beginning with the work of Altman (1968) and Beaver (1966). This is quite unsurprising as corporate failures are relevant to a wide array of stakeholders, including but not limited to employees, creditors, shareholders, and even customers and suppliers. In audit research the same topic is referred to as going-concern prediction, since auditors need to attest whether they believe that a firm will continue to operate in the following fiscal year. For simplicity, this chapter uses bankruptcy prediction as a general term. Similar to fraud detection, bankruptcy prediction is a binary classification problem of whether a company will declare bankruptcy. This line of research started with Beaver (1966), who was the first to employ financial statement variables to predict bankruptcy via logistic regression, and was advanced through the introduction of ML in the 1990s.

The early use of ML within bankruptcy prediction examined structured financial data such as profitability ratios, growth ratios and debt ratios (e.g., Boritz & Kennedy, 1995; Kumar & Ravi, 2007; J. Sun & Li, 2008). In addition to established and well-understood statistical learning methods such as logistic regression, many researchers now favor the supervised learning approach with neural networks, followed by SVMs (Lin et al., 2012). Hybrid models (i.e., the combination of different ML techniques), unsupervised learning techniques, and genetic algorithms[9] are also applied (Lin et al., 2012). In general, financial data significantly improves bankruptcy prediction accuracy; however, backward-looking financial statement data hardly contains all the relevant information to predict future bankruptcies. As a remedy, recent literature has introduced non-financial, unstructured textual data from annual reports (e.g., Hajek et al., 2014; Shirata et al., 2011), specifically from the MD&A section (e.g., Cecchini et al., 2010b). Features are constructed from this data, for example by counting words, extracting key phrases, or assigning textual sentiments. Although financial data still dominates the research on bankruptcy prediction, there is convincing evidence that the combination of financial and non-financial data results in superior predictive power (e.g., Cecchini et al., 2010b; du Jardin, 2016; Hajek et al., 2014). Similar to fraud detection research, deep learning is used to extract more meaningful information from unstructured and structured data to further improve predictive performance (e.g., Mai et al., 2019).

6.3.4 Other Fields

Thus far, all studies presented in this review use ML in a predictive sense. This does not always have to be the case, as shown by Bertomeu and colleagues (2021). The authors study conditional conservatism, which is an asymmetric stock market reaction to good versus bad news, whereby good news is associated with positive returns and bad news with negative returns. Traditionally, this has been done by reverse regression (i.e., regressing earnings on returns) with an interaction term between a negative return indicator and returns (Basu, 1997). It is shown that the coefficient of the interaction term is positive and significant, meaning that earnings react asymmetrically to good and bad news. Instead of simply identifying linearities in the identification of conservatism with a linear regression, the authors study the richness of non-linearities. Specifically, they treat the coefficients not as constants but as a functional form of other explanatory variables, such as book-to-market, size and leverage. These coefficients are fitted by a neural network and studied afterwards. The main objective is thus to understand the non-linear behavior of conservatism. They report three novel insights on conservatism: (1) conservatism has been decreasing since the 1980s and is now at an all-time low; (2) the relation between conservatism and size or book-to-market equity is convex, while it is concave between conservatism and leverage or volatility; and (3) the relation between conservatism, investment and cost of capital has a greater economic magnitude when using ML and less measurement noise.

6.4 Management Accounting

Management accounting, also called managerial accounting, is distinct from financial accounting and reporting in that it refers to internal reporting systems. The Institute of Management Accountants defines management accounting more specifically as “a profession that involves partnering in management decision making, devising planning and performance management systems, and providing expertise in financial reporting and control to assist management in the formulation and implementation of an organization’s strategy“ (Institute of Management Accountants, 2008). Accordingly, as compared to financial accounting, management accounting focuses more on value-creation, decision-making, planning and forecasting, as well as risk management (e.g., Wang & Wang, 2016). In practice, tasks of management accountants include cost management to achieve long-term goals and

---

[9] Genetic algorithms are inspired by natural selection, meaning a model evolves through generations and the allowance for random mutations, which will be kept if they serve the model objective.

objectives, management and operational control through performance measurement, and internal (cost) activity planning (e.g., budgeting). Use of companies' big data thus gives management accountants the perfect tools to answer questions about what has happened, what likely will happen, and what is the optimized solution. For these tasks, they can employ descriptive statistics, predictive analytics (i.e., extrapolating the status quo into the future), and prescriptive analytics (i.e., prescribing the optimal solution or suggest changes to operational practices) within each of these tasks (e.g., Appelbaum et al., 2017). ML is ideal for predictive analytics, which arguably is becoming more important than descriptive statistics because of the intensified competition triggered by globalization and technological advances (see Cokins, 2013). This section considers exemplary studies of performance management, cost management and risk control.

Performance management can be subdivided into employee performance evaluation and project performance evaluation. With employee performance evaluation, ML is used to support decision-making in the human resource management functions of hiring and monitoring processes (Hagemann and Klug, 2022). In terms of methodology, variants of decision trees are most often used due to their interpretability. For example, Al-Radaideh and Al Nagi (2012) study features which may affect employees' performance in IT companies with the help of two decision tree variations. The target (i.e., employee performance) consists of three classes: *accomplish*, *exceed* and *far exceed*. They report that the job title and the type of university from which the employee graduated have the strongest predictive power. Based on the same classification techniques, Kirimi and Moturi (2016) conduct a similar study with data from the Kenyan School of Government. Here, the employees' experience has the strongest predictive power. With respect to project performance evaluation, ML is used to analyze large amounts of data to extract hidden knowledge which helps decision-makers to manage workforces, cash flows and other resources. For more general purposes of company performance evaluation, Cheriyan and colleagues (2018) predict sales of e-commerce companies. They compare different models, with gradient-boosted regression trees exhibiting the best predictive performance. Another means to evaluate companies' performance is through analyzing costumer satisfaction. As an example, Katarya and colleagues (2020) train a model on Yelp[10] online reviews, with the text of reviews representing features and the star ratings assigned by customers representing sentiment. Every review with three or more stars is coded as one, and every other review as zero, i.e., it is a binary classification task. The trained model can then be applied to reviews without stars in order to automatically evaluate these costumers' sentiment towards the firm.

Cost management follows the cycle of resource planning, cost estimation, cost budgeting and cost control (Bhimani et al., 2019). An important task of management accountants is to correctly estimate costs of upcoming projects to support the firms' strategic decision-makers efficiently (Lum et al., 2008). It mainly takes place in the pre-project planning phase, as the vast majority of costs are only influencable early-on. Estimating construction projects is a prominent area in which ML is already integrated into cost management (Hashemi et al., 2020). Here, the aim of ML is to increase predictive power by taking advantage of non-linearities while using more features and simultaneously being able to handle multicollinearities. Son and colleagues (2012), for example, incorporate a hybrid model consisting of a PCA to deal with multicollinearities and an SVR to predict the cost performance of commercial building projects. They report higher predictive performance compared to other commonly used data

[10] Yelp is a crowd-sourced platform for business reviews.

mining techniques such as neural networks, decision trees and stand-alone SVRs. A more adapted neural network, with various regularization techniques for better generalization and less overfitting, is used by Chandanshive and Kambekar (2019), who also report successful predictive abilities. While the aforementioned studies only use structured data as features, Williams and Gong (2014) combine structured (financial) data with unstructured textual data. The authors predict construction cost overruns and show that certain words and phrases contained in tender (i.e., bidding) documents are associated with unplanned cost increases for construction projects. Other industries in which the use of ML to estimate costs is documented in the scientific literature are software development and healthcare (e.g., Al Asheeri & Hammad, 2019; Koh & Tan, 2011).

In the area of risk control, ML is used to detect early warning signs of financial distress that could potentially result in bankruptcy. This makes this line of research comparable to the going-concern-related research in the audit literature. Koyuncugil and Ozgulbas (2012) present such a system for small and medium-sized enterprises. Their focus is on building a system that is understandable for managers and decision-makers who do not have financial or data mining expertise. For that reason, they use a variation of the highly interpretable decision trees. With financial features (categorized into liquidity ratios, ratios of financial position, turnover ratios and profitability ratios) and binary targets (e.g., poor versus strong financial performance), the authors identify two meaningful early warning signs, namely profit before tax to own funds and return on equity. Credit scoring is another area in which ML supports the risk control function. Better decisions about whether loan applicants are eventually accepted result in reduced creditor risks and, most importantly, costs (Huang et al., 2007). In their study, Huang and colleagues (2007) use the combination of a genetic algorithm for feature selection and an SVM for prediciton purposes, i.e., a hybrid model. The SVM is used due to the high performance with relatively few features compared to neural network or decision tree classifiers. Since even small accuarcy improvements are effective in terms of future cost savings, it does not come as a great surprise that the use of data mining in this field is studied quite rigorously (M. R. Kumar & Gunjan, 2020). A recent trend in the credit scoring literature is the use of deep learning (e.g., Dastile & Celik, 2021; Gunnarsson et al., 2021; Wu et al., 2021).

In summary, ML in management accounting allows hidden relations and tendencies to be uncovered while using different data types such as semi-structured and unstructured data, or by using more data overall (Wang & Wang, 2016). We find that the use of ML is comparatively well established in the financial accounting and audit literature. In management accounting, ML is still a niche topic, apart from the literature on credit scoring, which shows extensive similarities to, for instance, the research on predicting bankruptcies.

6.5 State of the Art and Future Opportunities

In most fields, the application of ML has followed a similar path. Suitable research areas for ML initially evolved by using simple, structured data sets that were mostly processed using classical statistical models, such as logistic regressions for classification and linear regression for regression tasks. With the rise of ML, such tasks have been applied more often across many fields, but to similar research questions as before. With generally greater acceptance of ML, researchers in the field turn to methodological improvements, such as the comparison of individual ML techniques and model ensemble creations. As the usefulness of unstructured data becomes recognized, combined applications of structured and unstructured data follow suit. Deep learning is then introduced to extract information most efficiently from structured and unstructured data alike. A simultaneous phase focuses on interpretability with, for example, explainable ML, although its application nevertheless remains rare.

As can be seen in the fraud detection literature, structured and unstructured data are already studied, but deep learning is applied only partially. Future research could further concentrate on deep learning or on explainable ML to gain a better understanding of the relations between features and targets (Psychoula et al., 2021). More specifically, the time-series dependence of accounting fraud with various features could be studied. This could then be used to uncover still hidden fraud cases, providing auditors, regulators and investors with an early warning system for fraudulent behavior. Unsupervised learning or semi-supervised learning can be used for this purpose. Although the imbalance problem is solved by changing the sample composition through random sampling, it is also possible to apply classification models which are well suited for imbalanced datasets (see Kondo et al., 2019). Such models use a weighted random forest instead of a simple random forest. Another model warranting consideration would be the extreme gradient-boosting algorithm, which is insentive to imbalances (Zięba et al., 2016).

Similar to accounting fraud detection, the bankruptcy prediction literature already exploits the use of structured and unstructured data as well as the application of deep learning techniques. Despite extensive research using various data types and classification techniques, the predictive performance has not improved by much in recent years, and the investigation of predictive features is still called for (Lin et al., 2012; Tang et al., 2020). A possible solution would be a greater focus on model interpretability. Researchers could thereby answer hypotheses more efficiently, and practitioners' acceptance of ML-based bankruptcy prediction models would increase (H. Son et al., 2019). To our knowledge, there are not many studies with a focus on interpretability. The work of Tang and colleagues (2020) is among the rare exceptions. They compare different prediction horizons and many ML techniques, from simple logistic regression over SVMs to deep neural networks, with a focus on interpretability – not necessarily model interpretability, but rather the interplay between feature (data) types, prediction horizons and predictive performance. Methodically, they see great potential in the use of deep learning, which could be especially valuable given its capabilities in processing structured and unstructured data simultaneously.

Another improvement might entail the implicit handling of imbalances. Bankruptcies, much like accounting fraud, are rare events in real-world data, rarely exceededing 1 percent of the population (Kuger, 2019). Therefore, it is necessary to deal with the imbalance problem, which is typically ignored or bypassed by drawing evenly distributed samples. There are only some studies that deal with this imbalance problem implicitly, by, for example, modifying the sample distribution through random sampling or using ML algorithms which are suitable under an imbalanced dataset, e.g., extreme gradient-boosting tree (e.g., Gruszczyński, 2019; Le et al., 2018a; Le et al., 2018b; Zięba et al., 2016).

For management accounting research, specifically cost estimation and decision-making support, ML is particularly suited to processing external data, such as data from suppliers, partners and competitors (Wang & Wang, 2016). So far, mainly cost figures estimated on internal business operations have been used (Lanen et al., 2013). The application of deep learning would also be feasible since most external data are available in unstructured form. Lastly, for risk control there is great potential for predicting such outcomes as sales, cash flows and earnings to detect discontinuities early on, in the manner of an early warning system.

References

Abdolmohammadi, M. J. (1991). Factors affecting auditors' perceptions of applicable decision aids for various audit tasks. *Contemporary Accounting Research*, 7(2), 535–548.

Al-Radaideh, Q. A., & Al Nagi, E. (2012). Using data mining techniques to build a classification model for predicting employees performance. *International Journal of Advanced Computer Science and Applications*, 3(2), 144–151.

Al Asheeri, M. M., & Hammad, M. (2019). Machine learning models for software cost estimation. *International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies*.

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

Appelbaum, D., Kogan, A., Vasarhelyi, M. A., & Yan, Z. (2017). Impact of business analytics and enterprise systems on managerial accounting. *International Journal of Accounting Information Systems*, 25, 29–44.

Association of Certified Fraud Examiners (2020). *Report to the Nations: 2020 Global Study on Occupational Fraud and Abuse*.

Baldwin, A. A., Brown, C. E., & Trinkle, B. S. (2006). Opportunities for artificial intelligence development in the accounting domain: The case for auditing. *Intelligent Systems in Accounting, Finance and Management*, 14(3), 77–86.

Bao, Y., Hilary, G., & Ke, B. (2020a). Artificial intelligence and fraud detection. In *Innovative technology at the interface of finance and operations* (Working Paper).

Bao, Y., Ke, B., Li, B., Yu, J., & Zhang, J. (2020b). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. *Journal of Accounting Research*, 58(1), 199–235.

Basu, S. (1997). The conservatism principle and the asymmetric timeliness of earnings. *Journal of Accounting and Economics*, 24(1), 3–37.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111.

Bellman, R. E. (1961). *Adaptive control processes*. Princeton University Press.

Bertomeu, J., Cheynel, E., Floyd, E., & Pan, W. (2020). Using machine learning to detect misstatements. *Review of Accounting Studies*, 26, 468–519.

Bertomeu, J., Cheynel, E., Liao, Y., & Milone, M. (2021). Using machine learning to measure conservatism. *Financial Accounting EJournal*.

Bhimani, A., Datar, S. M., Horngren, C. T., & Rajan, M. V. (2019). *Management and cost accounting* (7th ed.). Pearson Education, Limited.

Boritz, E. J., & Kennedy, D. B. (1995). Effectiveness of neural network types for prediction of business failure. *Expert Systems with Applications*, 9(4), 503–512.

Brandas, C., Muntean, M., & Didraga, O. (2018). Intelligent decision support in auditing: Big data and machine learning Approach. *Proceedings of the IE 2018 International Conference*.

Brown, N. C., Crowley, R. M., & Elliott, W. B. (2020). What are you saying? Using topic to detect financial misreporting. *Journal of Accounting Research*, 58(1), 237–291.

Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. *Accounting Horizons*, 29(2), 423–429.

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010a). Detecting management fraud in public companies.

Management Science, 56(7), 1146–1160.

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

Chandanshive, V. B., & Kambekar, A. R. (2019). Estimation of building construction cost using artificial neural networks. *Journal of Soft Computing in Civil Engineering*, 3(1), 91–107.

Chen, C., Liaw, An., & Breiman, L. (2004). Using random forest to learn imbalanced data. In *Technical Report 666, Statistics Department of University of California at Berkley*.

Cheriyan, S., Ibrahim, S., Mohanan, S., & Treesa, S. (2018). Intelligent sales prediction using machine learning techniques. *International Conference on Computing, Electronics & Communications Engineering (ICCECE)*.

Cokins, G. (2013). Top 7 trends in management accounting. *Strategic Finance*, 95(6), 21–30.

Craja, P., Kim, A., & Lessmann, S. (2020). Deep learning for detecting financial statement fraud. *Decision Support Systems*, 139, 113421.

Dastile, X., & Celik, T. (2021). Making deep learning-based predictions for credit scoring explainable. *IEEE Access*, 9, 50426–50440.

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.

Drummond, C., & Holte, R. C. (2003). C4.5, class imbalance, and cost sensitivity: Why undersampling beats oversampling. *Proceedings of the Twentieth International Conference on Machine Learning: Workshop – Learning from Imbalanced Data Sets II.*, 1–8.

du Jardin, P. (2016). A two-stage classification technique for bankruptcy prediction. *European Journal of Operational Research*, 254(1), 236–252.

Dutta, I., Dutta, S., & Raahemi, B. (2017). Detecting financial restatements using data mining techniques. *Expert Systems with Applications*, 90(30), 374–393.

Earley, C. E. (2015). Data analytics in auditing: Opportunities and challenges. *Business Horizons*, 58(5), 493–500.

Fanning, K. M., & Cogger, K. O. (1998). Neural network detection of management fraud using published financial data. *Intelligent Systems in Accounting, Finance and Management*, 7(1), 21–41.

Gepp, A., Linnenluecke, M. K., O'Neill, T. J., & Smith, T. (2018). Big data techniques in auditing research and practice: Current trends and future opportunities. *Journal of Accounting Literature*, 40, 102–115.

Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow*. O'Reilly Media, Inc.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Green, B. P., & Choi, J. H. (1997). Assessing the risk of management fraud through neural network technology. *Auditing: A Journal of Practice & Theory*, 16(1), 14–28.

Gruszczyński, M. (2019). On unbalanced sampling in bankruptcy prediction. *International Journal of Financial Studies*, 7(2), 28.

Gunnarsson, B. R., vanden Broucke, S., Baesens, B., Óskarsdóttir, M., & Lemahieu, W. (2021). Deep learning for credit scoring: Do or don't? *European Journal of Operational Research*, 295(1), 292–305.

Hagemann, V., & Klug, K. (2022). Human resource management in a digital environment. In L. Hornuf (Ed.), *Diginomics Research Perspectives: The Role of Digitalization in Business and Society*, (pp. @@@). Cham:

Springer International Publishing.

Hajek, P., Olej, V., & Myskova, R. (2014). Forecasting corporate financial performance using sentiment in annual reports for stakeholders' decision-making. *Technological and Economical Development of Economy*, 20(4), 721–738.

Hashemi, S. T., Ebadati, O. M., & Kaur, H. (2020). Cost estimation and prediction in construction projects: A systematic review on machine learning techniques. *SN Applied Science*, 1703.

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

Hoitash, R., & Hoitash, U. (2018). Measuring accounting reporting complexity with XBRL. *The Accounting Review*, 93(1), 259–287. https://doi.org/10.2308/ACCR-51762

Hooda, N., Bawa, S., & Rana, P. S. (2018). Fraudulent firm classification: A case study of an external audit. *Applied Artificial Intelligence*, 32(1), 48–64.

Huang, C.-L., Chen, M.-C., & Wang, C.-J. (2007). Credit scoring with a data mining approach based on support vector machines. *Expert Systems with Applications*, 33(4), 847–856.

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.

Institute of Management Accountants (2008). *Statements on management accounting: Definition of management accounting*. www.imanet.org

International Auditing and Assurance Standards Board (2016). *Exploring the growing use of technology in the audit, with a focus on data analytics*. New York, NY: IFAC.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer. https://doi.org/10.1007/978-1-4614-7138-7

Kashyap, P. (2017). *Machine learning for decision makers*. Apress.

Katarya, R., Gautam, A., Bandgar, S. P., & Koli, D. (2020). Analyzing customer sentiments using machine learning techniques to improve business performance. *International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*.

Kinra, A., Kotzab, H., & Siekmann, F. (2022). Social media analytics in operations and supply chain management: Opportunities, challenges and paradoxes. In L. Hornuf (Ed.), *Diginomics Research Perspectives: The Role of Digitalization in Business and Society*, (pp. @@@). Cham: Springer International Publishing.

Kirimi, J. M., & Moturi, C. A. (2016). Application of data mining classification in employee performance prediction. *International Journal of Computer Applications*, 146(7), 28–35.

Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. *Journal of Healthcare Information Management*, 19(2), 64–72.

Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. *Journal of Emerging Technologies in Accounting*, 14(1), 115–122.

Kondo, S., Miyakawa, D., Shiraki, K., Suga, M., & Usuki, M. (2019). Using machine learning to detect and forecast accounting fraud. *RIETI*.

Kotsiantis, S., Koumanakos, E., Tzelepis, D., & Tampakas, V. (2006). Forecasting fraudulent financial statements using data mining. *International Journal of Computational Intelligence*, 3(2), 104–110.

Koyuncugil, A. S., & Ozgulbas, N. (2012). Financial early warning system model and data mining application for risk detection. *Expert Systems with Applications*, 39(6), 6238–6253.

Kuger, M. (2019). *Global bankruptcy report 2019*. Dun & Bradstreet Worldwide Network.

Kumar, M. R., & Gunjan, V. K. (2020). Review of machine learning models for credit scoring analysis. *Ingeniería Solidaria*, 16(1).

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques: A review. *European Journal of Operational Research*, 180(1), 1–28.

Lanen, W., Anderson, S., & Maher, M. (2013). *Fundamentals of cost accounting* (4th ed.). McGraw-Hill Education.

Le, T., Lee, M. Y., Park, J. R., & Baik, S. W. (2018a). Oversampling techniques for bankruptcy prediction: Novel features from a transaction dataset. *Symmetry*, 10(4), 79.

Le, T., Son, L. H., Vo, M. T., Lee, M. Y., & Ba0ik, S. W. (2018b). A cluster-based boosting algorithm for bankruptcy prediction in a highly imbalanced dataset. *Symmetry*, 10(7), 250.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293–303.

Li, B., Yu, J., Zhang, J., & Ke, B. (2016). Detecting accounting frauds in publicly traded U.S. firms: A machine learning approach. *Asian Conference on Machine Learning, PMLR 45*, 173–188.

Lin, W.-Y., Hu, Y.-H., & Tsai, C.-F. (2012). Machine learning in financial crisis prediction: A survey. *IEEE Transactions on Systems, Man, and Cybernetics*, 42(4), 421–436.

Lum, K. T., Baker, D. R., & Hihn, J. M. (2008). The effects of data mining techniques on software cost estimation. *IEEE International Conference on Engineering Management (EMC)*.

Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274(2), 743–758.

Melo-Acosta, G. E., Duitama-Munoz, F., & Arias-Londono, J. D. (2017). Fraud detection in big data using supervised and semi-supervised learning techniques. *Proceedings of the 2017 IEEE Colombian Conference on Communications and Computing, COLCOM 2017*. https://doi.org/10.1109/COLCOMCON.2017.8088206

Mishra, B. K., Hazra, D., Tarannum, K., & Kumar, M. (2016). Business intelligence using data mining techniques and business analytics. *International Conference on System Modeling & Advancement in Research Trends (SMART)*.

No, W. G., Lee, K. K., Huang, F., & Li, Q. (2019). Multidimensional audit data selection (MADS): A framework for using data analytics in the audit data selection process. *Accounting Horizons*, 33(3), 127–140.

Office of Compliance Inspections and Examinations of the Securities and Exchange Commission (2015). *Examination priorities for 2015*. http://www.sec.gov/about/offices/ocie/national-examination-program-priorities-2015.pdf

Perols, J. L. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice & Theory*, 30(2), 19–50.

Perols, J. L., Bowen, R. M., Zimmermann, C., & Samba, B. (2017). Finding needles in a haystack: Using data analytics to improve fraud prediction. *The Accounting Review*, 92(2), 221–245.

Psychoula, I., Gutmann, A., Mainali, P., Lee, S. H., Dunphy, P., & Petitcolas, F. A. P. (2021). Explainable machine learning for fraud detection. *IEEE Computer Special Issue on Explainable AI and Machine Learning*, forthcoming.

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.

Shirata, C. Y., Takeuchi, H., Ogino, S., & Watanabe, H. (2011). Extracting key phrases as predictors of corporate bankruptcy: Empirical analysis of annual reports by text mining. *Journal of Emerging Technologies in Accounting*, 8(1), 31–44.

Son, H., Hyun, C., Phan, D., & Hwang, H. J. (2019). Data analytic approach for bankruptcy prediction. *Expert Systems with Applications*, 138(30), 112816.

Son, Hyojoo, Kim, C., & Kim, C. (2012). Hybrid principal component analysis and support vector machine model for predicting the cost performance of commercial building projects using pre-project planning variables. *Automation in Construction*, 27, 60–66.

Sun, J., & Li, H. (2008). Data mining method for listed companies' financial distress prediction. *Knowledge-Based Systems*, 21(1), 1–5.

Sun, T. (2019). Applying deep learning to audit procedures: An illustrative framework. *Accounting Horizons*, 33(3), 89–109.

Sun, T., & Vasarhelyi, M. A. (2017). Detailed record title: Deep learning and the future of auditing: How an evolving technology could transform analysis and improve judgment. *CPA Journal*, 87(6), 25–29.

Tang, X., Li, S., Tan, M., & Shi, W. (2020). Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods. *Journal of Forecasting*, 39(5), 769–787.

Wang, Y., & Wang, Z. (2016). Integrating data mining into managerial accounting system: Challenges and opportunities. *Chinese Business Review*, 15(1), 33–41.

Williams, T. P., & Gong, J. (2014). Predicting construction cost overruns using text mining, numerical data and ensemble classifiers. *Automation in Construction*, 43, 23–29.

Wu, C.-F., Huang, S.-C., Chiou, C.-C., & Wang, Y.-M. (2021). A predictive intelligence system of credit scoring based on deep multiple kernel learning. *Applied Soft Computing*, 111, 107668.

Zięba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications*, 58, 93–101.