

Identifying the limitations associated with machine learning techniques in performing accounting tasks

Machine
learning
techniques

227

Liezl Smith and Christiaan Lamprecht
School of Accountancy, Stellenbosch University, Stellenbosch, South Africa

Received 30 May 2023
Revised 4 October 2023
14 December 2023
23 January 2024
20 February 2024
Accepted 7 March 2024

Abstract

Purpose – In a virtual interconnected digital space, the metaverse encompasses various virtual environments where people can interact, including engaging in business activities. Machine learning (ML) is a strategic technology that enables digital transformation to the metaverse, and it is becoming a more prevalent driver of business performance and reporting on performance. However, ML has limitations, and using the technology in business processes, such as accounting, poses a technology governance failure risk. To address this risk, decision makers and those tasked to govern these technologies must understand where the technology fits into the business process and consider its limitations to enable a governed transition to the metaverse. Using selected accounting processes, this study aims to describe the limitations that ML techniques pose to ensure the quality of financial information.

Design/methodology/approach – A grounded theory literature review method, consisting of five iterative stages, was used to identify the accounting tasks that ML could perform in the respective accounting processes, describe the ML techniques that could be applied to each accounting task and identify the limitations associated with the individual techniques.

Findings – This study finds that limitations such as data availability and training time may impact the quality of the financial information and that ML techniques and their limitations must be clearly understood when developing and implementing technology governance measures.

Originality/value – The study contributes to the growing literature on enterprise information and technology management and governance. In this study, the authors integrated current ML knowledge into an accounting context. As accounting is a pervasive aspect of business, the insights from this study will benefit decision makers and those tasked to govern these technologies to understand how some processes are more likely to be affected by certain limitations and how this may impact the accounting objectives. It will also benefit those users hoping to exploit the advantages of ML in their accounting processes while understanding the specific technology limitations on an accounting task level.

Keywords Accounting tasks, Artificial intelligence, Digital transformation, Grounded theory, Technology governance, Machine learning, Metaverse

Paper type Research paper

Introduction

The recent COVID-19 pandemic led many businesses to undergo accelerated digital transformation (Deloitte, 2020; Lee *et al.*, 2021; World Economic Forum, 2021), resulting in



© Liezl Smith and Christiaan Lamprecht. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Journal of Financial Reporting and
Accounting
Vol. 22 No. 2, 2024
pp. 227-253
Emerald Publishing Limited
1985-2517
DOI 10.1108/JFRA-05-2023-0280

significantly faster implementation of long-term strategic technology plans than intended. This, in turn, may have increased the risk of a technology governance failure, one of the critical business risks identified by the Global Risk Report (World Economic Forum, 2021).

One of the aims of digital transformation is to benefit the way an organisation operates through, amongst other things, digital technologies. One example is enabling automation using artificial intelligence (AI) (Vial, 2019). According to Mardini and Alkurdi (2021), AI will automate most financial accounting tasks in the future. Machine learning (ML), a subset of AI, saw a massive increase in the development of tools and applications since 2019, with one of the top use cases being process automation (Algorithmia, 2020), such as capturing documents, classifying transactions, account reconciliations and preparation of financial reports.

The maturation of ML and AI is also one of the technologies enabling digital transformation to the metaverse (Lee *et al.*, 2021; Blau *et al.*, 2022), a virtual world where people can interact in various activities, including engaging in business activities that require accounting (Pandey and Gilmour, 2023). Commentators argue that the metaverse will be the culmination of various technologies to create an immersive platform which extends the way business is currently performed from the physical world to the digital world (Lee *et al.*, 2021; Blau *et al.*, 2022). Blau *et al.* (2022) identify four key factors that will impact the metaverse's potential, one of which is governance.

The risks of technology governance failure, the impact of governance on digital transformation to a metaverse and the uncertainty surrounding how this transformation will achieve desired outcomes emphasise the need for adequate technology governance to ensure that digital transformation does, in fact, advance and achieve the objectives of the businesses, reducing the risk of technology governance failure. The risk of technology governance failure surrounding implementing new technologies, such as ML, prompted the King IV Report to require a business to govern technology to achieve its objectives (Institute of Directors of Southern Africa [IODSA], 2016). Alreemy *et al.* (2016) concur in their description of the aim of technology governance, namely, to ensure compatibility between the goals of the business and a satisfactory level of risk with the use of emerging technologies.

There are two aspects to technology governance: governing business objectives using technology and the actual governance of technology to ensure the technology achieves its objectives (Wilkin and Chenhall, 2020). With a focus on the latter, obtaining buy-in from stakeholders and alignment within the organisation between various stakeholders were some of the challenges found when implementing ML tools (Algorithmia, 2020). Technology governance can be applied at a strategic and operational level (Goosen and Rudman, 2013). Our research focuses on strategic technology governance at a business process level, assisting accounting users to ensure business information technology (IT) alignment, as Goosen and Rudman (2013) recommended. The accounting process, one of the business processes most prevalent in organisations, can benefit from ML. ML enables the automation of many tasks in the accounting process, reducing the risk of human error and making these tasks more efficient (Fallatah, 2021). Tasks may include processing source documents and analysing business transactions (Cho *et al.*, 2020).

Considering the link between achieving business objectives, technology governance and the digital technologies used, such as AI and ML, and in the context of accounting tasks, it would follow that it is necessary to understand what the ML technology should achieve and then to understand its limitations, that is, what would prevent the technology from reaching the stated objectives. Bavaresco *et al.* (2023) and Kommunuri (2022) identify the importance of understanding the technology's limitations. In this way, the technology can be governed to reduce the risk of stated objectives not being achieved, as King IV requires. In considering the general limitations of ML that impact the accounting process, existing research focuses

on the general limitations of using ML technology, such as the lack of interpretability of algorithms and algorithmic bias (Cho *et al.*, 2020; Fallatah, 2021; PWC, 2019). However, to better understand the limitations of specific ML techniques in the accounting context, accounting decision makers need to know where ML fits into each accounting process and the specific limitations that may consequently arise. The problem is that the limitations of those ML techniques that can perform particular accounting tasks within the accounting process are not identified and, therefore, cannot be adequately considered and governed. An accounting task can be defined as any action to record an economic event, adjustment or modification according to accounting principles (Petkov, 2020).

This research aims to identify the limitations of some ML techniques that can perform specific accounting tasks within the accounting process. To achieve the research aim and answer the research problem, the authors formulated the following research questions to guide the research process:

- RQ1.* Which tasks in the accounting processes can be assisted or performed by ML techniques?
- RQ2.* What are the limitations associated with the identified ML techniques, and do these link to the accounting objectives?

This study finds that limitations such as data availability and training time may impact the quality of the financial information and that ML techniques and their limitations must be clearly understood when developing and implementing technology governance measures. This paper contributes to the field of technology governance research, specifically considering the limitations of using ML in accounting against the qualitative objectives of useful accounting information. Moreover, the study will benefit those accounting users hoping to exploit the advantages of ML in their accounting processes while understanding the specific technology limitations on an accounting task level. It will assist especially those accounting decision makers wanting to know how some processes are more likely to be affected by certain limitations and how this may impact the achievement of business objectives, specifically the qualitative accounting objectives.

The remainder of the article is laid out as follows: The next section presents a brief literature review to explain the aspects impacting the research problem, followed by the research design. Next, the analysis and findings are presented, followed by conclusions based on our research findings. The article closes with the limitations and suggestions for future research.

Literature review

Accounting objectives and quality of financial information

Gillion (2017) states that in all businesses, accounting processes aim to produce high-quality accounting information for decision-making and, therefore, high-quality financial reports. The *Conceptual Framework for Financial Reporting* describes and categorises the qualitative characteristics of useful financial information (International Accounting Standards Board, 2022) into fundamental and enhancing characteristics. For this study, the fundamental and enhancing qualitative characteristics are designated as the qualitative accounting objectives. Later, we will link any applicable limitations of a particular ML technique used to produce useful financial information to these qualitative accounting objectives. We briefly define and summarise the accounting objectives in Table 1.

This section has listed the qualitative accounting objectives of useful financial information. The tasks in these accounting processes will be set out next.

Qualitative accounting objective	Description
<i>Fundamental</i>	
Relevance	Information needs to be relevant to the decisions users are making. Information influences decisions if it can be used to predict future outcomes or to confirm prior evaluations
Faithful representation	The information must represent the substance of the presented matter, not just the form. To do this, the information should be complete, neutral and error-free
Materiality	Information is material if ignoring it or misstating it could affect decisions
<i>Enhancing</i>	
Comparability	Accounting information needs to be comparable and enable users to identify the similarities and differences in information. Consistency helps to achieve this goal
Verifiability	The information needs to be verified in some way, either by direct observation or by recalculating the outputs using the known inputs and methods used
Timeliness	Information needs to be available to users in time to make the required decisions
Understandability	Information needs to be presented and classified clearly and concisely
<i>Pervasive (authors' classification)</i>	
Cost vs benefit (cost-saving)	The cost constraint of useful financial information, as opposed to the benefits to the user

Table 1.
Qualitative
accounting objectives

Source: Compiled by author from [International Accounting Standards Board \(2022\)](#)

Accounting process and tasks to produce financial information

In the traditional accounting process, source documents received and generated are captured in an accounting record system, reconciled, and finally, the financial information produced is presented in financial reports. In this study, we used [Deming's \(2024\)](#) description of the traditional record-to-report process to identify the three main accounting processes. Each main process is then broken up into tasks to enable us to identify ML techniques that could perform those task(s). [Figure 1](#) illustrates Deming's traditional record-to-report process and our summation into the three main accounting processes (processes 1 to 3).

[Figure 1](#) illustrates how the record-to-report process commences with the external information sources (Process 1), followed by account reconciliations (Process 2) and then journal entries, month-end closure, analysis and reporting (Process 3). Our study does not address the performance of compliance and control procedures.

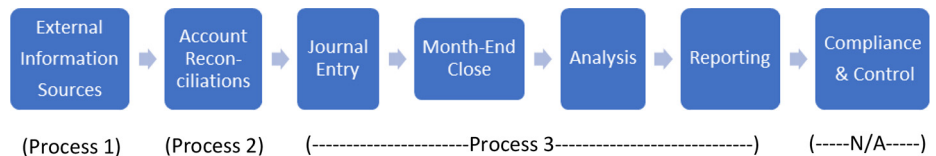


Figure 1.
Traditional record-to-report process

Source: Compiled by author from [Deming \(2024\)](#)

To identify one or more ML techniques to perform the accounting process, the accounting processes need to be broken down into tasks within that broad process. According to [Amani and Fadlalla \(2017\)](#), there is a paucity of published research on applications that use ML techniques. While the reason for this is unclear, they speculate that it may be due to a lack of reporting on such applications because of the unwillingness to reveal details of these applications for competitive reasons ([Amani and Fadlalla, 2017](#)). Based on available research, we have identified the following tasks within the respective three accounting processes and summarised these in [Figure 2](#).

[Figure 2](#) describes the various accounting tasks in the accounting processes. In the next section, we explore ML and the techniques that could perform some accounting tasks.

Machine learning technology to automate tasks

As a technology, ML is a subset of AI, in which data patterns are learned and applied in a changing environment. The technology does not require all possible situations to be known during development ([Ayodele, 2010a; Sainani, 2014](#)) and is one of the technologies that can be used in the accounting process to assist in the automation of tasks ([Everest Group, 2014](#)).

Despite the uncertainty, ML technology can detect patterns or predict solutions ([Valavanis et al., 1994](#)). To detect patterns or predict solutions, ML uses algorithms. This feature enables the technology to automate routine tasks, making it so valuable for automating routine accounting tasks ([Sapp, 2017](#)).

Based on the objective of the algorithm, how the algorithm learns, as well as the structure and volume of the data used for learning, ML algorithms can be distinguished as supervised, unsupervised and semi-supervised learning ([Ayodele, 2010b; Castle, 2018](#)). For the reader's benefit, the three categories are briefly described, as all three categories may have useful ML algorithms to automate accounting tasks.

Firstly, supervised learning algorithms require training. The algorithm is trained using a labelled data set consisting of examples of input data and labels indicating predicted targets or output data. Labels assist the algorithm in determining which features are essential. The

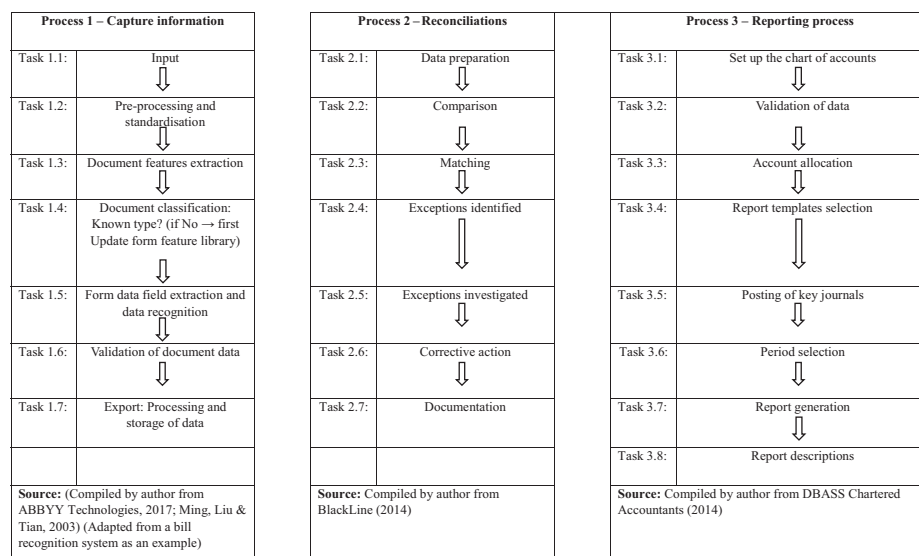


Figure 2.
Accounting tasks

algorithm then generalises the training set by mapping the inputs to the correct responses, which enables it to produce output for new inputs (Marsland, 2009; Ayodele, 2010b; Larsson and Segerås, 2016; Castle, 2018). The algorithm’s training on previous data sets, for example, the classification of accounting invoices and the respective financial fields, makes this type of ML particularly useful for accounting task automation, mainly due to the vast amount of data available and the rules-based nature of accounting.

Secondly, unsupervised learning algorithms do not require training. The input data are unlabelled, meaning the predicted values are not provided, which may be because they are unknown. The algorithm needs to determine the links between the inputs provided to identify patterns or commonalities that can be used to categorise new data or solve problems (Marsland, 2009; Ayodele, 2010b; Larsson and Segerås, 2016). An unsupervised algorithm is suitable where the action required, or outcome is uncertain for the task being performed or where unknown data needs to be grouped. This type of algorithm can be useful for automating accounting tasks such as error detection.

Thirdly, semi-supervised learning algorithms are trained using a combination of labelled and unlabelled data to generate an appropriate function. The labelled portion indicates patterns which may exist, while the unlabelled data, usually the larger portion of the data, are used to establish perceived or unknown patterns for the data (Ayodele, 2010b; Castle, 2018). It follows that where automation requires trained and untrained algorithms, such as allocating transactions to respective accounts, most transactions can be allocated using the prior knowledge from the training data. In contrast, a few transactions may not be known from the prior data, so a semi-supervised learning algorithm is preferable.

Machine learning techniques to solve problems

According to Someren and Urbancic (2006), the process of matching an identified problem to the technique to solve it is difficult. Firstly, the task must be understood. Secondly, a problem that ML can solve in that task must be identified and described. Thirdly, identifying the learning problem enables a developer to identify a solution, namely, the information and algorithm required to address the problem (Saitta and Neri, 1998; Someren and Urbancic, 2006).

Understanding the task and defining the learning problem is crucial, as many ML solutions are often available for addressing a problem (Someren and Urbancic, 2006). The types identified and described by Amani and Fadlalla (2017) and Larsson and Segerås (2016) are summarised in Table 2.

Table 2 shows the various ML solutions available to address an identified problem. One of the ML solutions is Classification (refer to Table 2). Kotsiantis (2007) explains that classification algorithms are supervised learners, and therefore, their development consists

ML solution	Description of use
Classification	Suitable for mapping data into two or more categories, each with its distinct attributes
Prediction	Suitable for producing a forward-looking numerical prediction (forecasting) or non-numerical prediction (classification)
Clustering	Suitable for separating data into classes or groups that are similar in some meaningful way
Outlier detection	Suitable for finding the items or events that significantly deviate from the expected pattern or other data considered normal in the data set

Table 2.
ML solutions

Source: Compiled by author from Larsson and Segerås (2016) and Amani and Fadlalla (2017)

of a two-step training and testing process. As indicated earlier, supervised learning algorithms are particularly useful for accounting task automation.

According to Kotsiantis (2007), the different supervised classification algorithms can further be separated into three types of techniques:

- (1) logic-based;
- (2) perceptron-based; and
- (3) statistical (Kotsiantis, 2007).

Using the three different types, Kotsiantis (2007) gives examples of the various ML algorithms for each type of supervised classification technique. It is these algorithms that can be used to solve a problem.

To expose the reader to the possible ML algorithms, we use the classification of ML techniques adapted from Kotsiantis (2007) to identify several algorithms. Figure 3 illustrates the type of ML techniques grouped according to the layout in Table 2.

It is outside the scope of this article to consider the detailed technical features of each of the algorithms in Figure 3. However, we point the reader to Appendix for a brief description of the general characteristics of the algorithms.

Given the increasing reliance on ML within financial systems, algorithmic bias, even small, could scale into substantial errors. For the benefit of the reader and practical relevance of the study we acknowledge these concerns and risks in the following section:

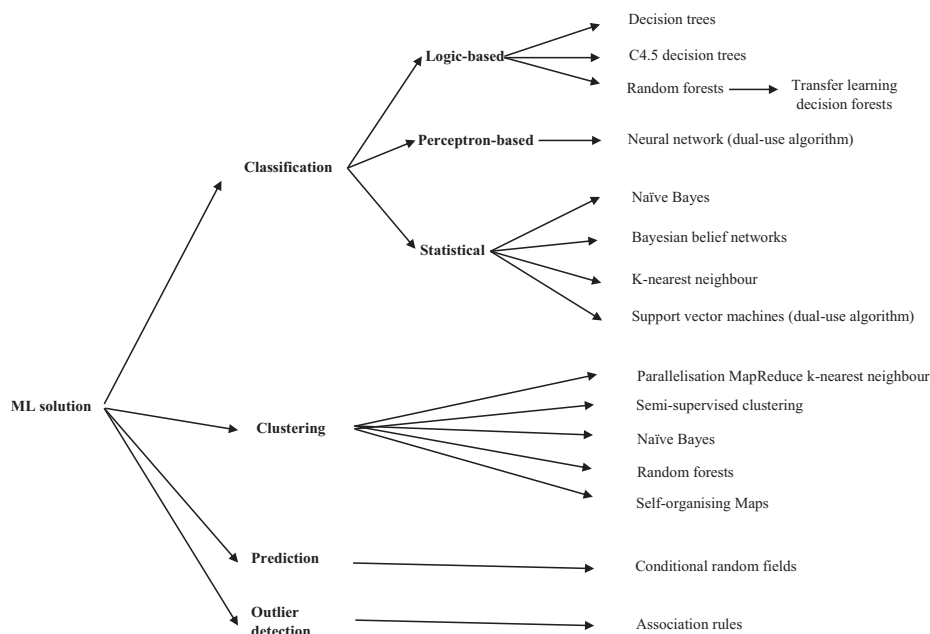


Figure 3. Types of ML techniques

Source: Created by author from multiple sources – refer to Appendix

Machine learning – general algorithmic bias

The increasing integration of ML in areas such as health care, education, employment and law raises the concern that unintended algorithmic biases can lead to adverse consequences (Barocas and Selbst, 2016; Kleinberg *et al.*, 2016; Mitchell *et al.*, 2019). In accounting, the general limitation of algorithmic bias, as highlighted in the research by Cho *et al.* (2020), is an important area that merits further exploration, albeit beyond the scope of this article. However, it is important for accounting users to be aware of this general limitation due to the associated risks of errors.

One illustrative example is in predicting accounting fraud, where Suresh and Gutttag (2019) have identified different types of bias that may be embedded in ML models, along with potential solutions. Notably, many of these solutions focus on addressing the quality of the data used, aligning with Gillion's assertion in 2017 that a successful machine-learning model requires high-quality data (Gillion, 2017).

Nevertheless, for a comprehensive overview of the types of general algorithmic bias and potential solutions, readers are directed to the excellent articles by Caton and Haas (2023), Cho *et al.* (2020) and Mehrabi *et al.* (2021). In this context, diverse and representative data sets with supporting documents (such as data sheets), labelling data, implementing statistical significance tests to detect discrimination and a commitment to ethical AI practices remain critical mitigation strategies to enhance fairness and reliability in ML applications.

The following section will explain how we analysed the phenomena to build a theory on the limitations of ML techniques that can perform specific accounting tasks within the accounting process.

Research design

Earlier, we noted that our research aims to identify the limitations of using ML technology for specific accounting tasks performed within the accounting process. To achieve this aim, we formulated research questions to focus on gaining theoretical insights into the limitations of ML techniques in an accounting process context. Therefore, the research is exploratory in nature.

An exploratory research design allows for the development of a grounded picture of the phenomena and the development of tentative theories or hypotheses (USC Libraries, 2024). Therefore, grounded theory (GT) will be used to achieve the research aim. GT is an inductive methodology used to develop a theory (Sutton *et al.*, 2011; Wolfswinkel *et al.*, 2013), and in this case, a theory about emergent technologies where there is not, as yet, an established theory. The GT method is particularly suited to information systems technology research (Bryant, 2002; Fernandez and Lehmann, 2011) and highlights the mandate of research to develop both an understanding of discovered facts and adequate models for specified purposes.

This study uses prior literature to identify the limitations of using ML technology for specific accounting tasks. We follow Wolfswinkel *et al.*'s (2013) five-stage process when conducting a GT literature review. The five stages are as follows:

- (1) define the scope of the review;
- (2) search the literature within the scope;
- (3) select the sample of literature to be analysed;
- (4) analyse the literature; and
- (5) present the results (Wolfswinkel *et al.*, 2013).

As this study aims to identify links between new variables, as is the purpose of a GT literature review method (Wolfswinkel *et al.*, 2013), a systematic approach was followed for each of the required variables. These variables are as follows:

- the *accounting tasks* in the selected accounting processes;
- the *ML techniques* available to perform the accounting tasks;
- the *limitations* of the ML techniques which were identified; and
- the *accounting objectives*.

The different stages of the GT literature review method were executed as follows.

Stage 1: define

As discussed earlier, the field of research is relevant to accounting practice, including information systems technology, financial processes and automation. Therefore, we considered literature on ML within an accounting context. However, if there were applicable examples and findings from practice, these were included in the scope of the research.

The initial search was broad and targeted to online databases. The search terms included keywords such as “machine learning,” “artificial intelligence,” “algorithm,” “accounting,” “financial,” “source document,” “invoice,” “reconciliation,” “reporting” and “automation” on the Scopus, EBSCOhost, IEEE and AAA digital library databases as recommended by Sutton *et al.* (2016). Following that, the search was expanded to websites and resources offered by accounting software providers (e.g. SAP and XERO), as we found that academic literature on accounting tasks was limited.

The final variable for inclusion was to identify the limitations of ML from existing literature. Here, the focus was first on the accounting literature and then expanded to information systems literature using the IEEE Xplore database. Search terms were based on the identified ML techniques combined with keywords in the search such as “risks,” “disadvantages” and “limitations.”

Stage 2: search

Having prepared the criteria and selected the appropriate sources and search terms, the searches were performed systematically. The researchers ensured that essential synonyms of search terms were included (Wolfswinkel *et al.*, 2013) if identified.

Stage 3: select

The literature’s abstracts identified in the search were read to determine if they were relevant to the aim of the study. In certain instances, the search criteria needed to be refined to find the relevant literature (Wolfswinkel *et al.*, 2013). Those papers that were then found to be in line with the aim of this research were downloaded for analysis.

Stage 4: analyse

The three selected accounting processes were investigated as a starting point to determine which tasks can be performed using ML techniques. To do this, each accounting process has been presented by separating it into its respective accounting tasks. The literature was then evaluated to identify which tasks offer learning problems: the task the ML should perform (Someren and Urbancic, 2006). Following that, the different types of ML techniques available to address each learning problem were identified and linked to the applicable accounting task.

Finally, the literature was reviewed to provide a list of the limitations of each ML technique. These limitations were considered in the context of the respective accounting tasks and linked to the applicable accounting process objectives. This link is made by considering the description of each qualitative accounting objective. This paper presents the Analysis (Stage 4) and Findings (Stage 5) together.

Stage 5: presentation

The last stage of the GT literature review method is to present the findings and insights gained in a structured manner. The following section presents the analysis (stage 4) and the findings (stage 5). The analysis and findings are linked and presented according to the particular research question.

Analysis and presentation of findings

Findings related to RQ1: which tasks in the accounting processes can be assisted or performed by machine learning techniques?

We discussed earlier that the process of matching each task to a technique involves, firstly, understanding the task, secondly, defining the learning problem and finally, identifying the information and algorithm required to address the problem (Saitta and Neri, 1998:137; Someren and Urbancic, 2006:366).

To match an identified learning problem to the technique to solve it, we explored each task in the three accounting processes to find whether a suitable ML technique can be applied to perform the identified task. We found and noted six learning problems in Process 1 (tasks 1.3, task 1.4 and task 1.6), one learning problem in Process 2 (task 2.3) and three learning problems in Process 3 (tasks 3.3, 3.7 and 3.8). We then consulted existing literature to find a suitable ML solution and type of ML technique(s) to address these learning problems. In Table 3, we summarise the learning problem identified per accounting process and task, and for each learning problem, the solution and the specific technique that achieves the solution.

The findings in Table 3 demonstrate that there may be more than one ML technique available to address a specific learning problem (Someren and Urbancic, 2006), and it is critical to understand the accounting process, the tasks within the process, the potential learning problems and what ML techniques can solve those problems. As there may be different limitations for each specific technique, it also follows that these need to be placed into the context of each accounting process, depending on the objectives.

Findings related to RQ2: what are the limitations associated with the identified machine learning techniques, and do these link to the accounting objectives?

To ensure adequate technology governance, as described by King IV (IODSA, 2016), the accountant would need to determine, for example, whether accuracy is more important than cost saving. Therefore, we considered the limitations of each ML technique that may impact the respective qualitative accounting objectives and are identified. We present our findings in Table 4.

Having considered the respective limitations of each ML technique we identified from the existing literature, Table 5 links these limitations to (an) applicable qualitative accounting objective(s). Moreover, we table the applicable tasks in the accounting process for each ML technique for the reader's benefit.

Description of the learning problem	Solutions to learning problem	ML techniques	Source
<i>Process 1: Translation of manual and electronic documents into accounting information</i>			
<i>Task 1.3 Document features extraction</i>			
<i>Feature extraction</i> is an important process of obtaining relevant data before the classification of images. This process can be improved using ML to perform deep feature extraction	Classification	<ul style="list-style-type: none"> Deep convolutional neural network 	Goussies, Ubalde, Fernandez, <i>et al.</i> (2014), Tarawneh <i>et al.</i> (2019)
<i>Task 1.4 Document type recognition and classification</i>	Classification	<ul style="list-style-type: none"> Convolutional neural networks New document class: k-nearest neighbour Similar known documents: support vector machine 	ABBYY Technologies (2017), Oquab, Bottou, Laptev, <i>et al.</i> (2014), Sorio (2013), Sorio, Bartoli, Davanzo, <i>et al.</i> (2010), Witten, Frank, Hall, <i>et al.</i> (2016), Khan (2019), Tarawneh <i>et al.</i> (2019), Chen <i>et al.</i> (2015)
<i>Irregular document layout classification using NLP</i> combined with ML to train the system to process flexible or irregular document layouts	Classification	Convolutional neural networks	Zhang <i>et al.</i> (2015), ABBYY Technologies (2017), Du (2017), Desai <i>et al.</i> (2021)
<i>Text classification</i> is used to classify text using statistical and semantic text analysis	Classification and Clustering	<i>Classification:</i> <ul style="list-style-type: none"> Naïve Bayes <i>Clustering:</i> <ul style="list-style-type: none"> Parallelisation MapReduce k-nearest neighbour Semi-supervised clustering 	

(continued)

Table 3. Accounting problem types and available ML techniques

Description of the learning problem	Solutions to learning problem	ML techniques	Source
<p><i>Task 1.6 Validation of document data:</i> <i>Validation of document information</i> can use ML to determine whether the extracted data from the document is correctly classified</p> <p><i>Removing duplicate entries and linking documents may be achieved using approximate string matching and ML for string classification</i></p>	<p>Classification</p> <p>Classification</p>	<ul style="list-style-type: none"> • Naïve Bayes • Support vector machine • Naïve Bayes • Decision trees • Support vector machine • Artificial neural network 	<p>Larsson and Segerås (2016)</p> <p>Antrup, Thompson, Kilby, <i>et al.</i> (2015), Larsson and Segerås (2016), De Leone and Minnetti (2015), Samoil (2015), Winkler (2014)</p>
<p><i>Process 2: Reconciliation of financial information</i> <i>Task 2.3 Matching</i> <i>Matching records or record-linkage</i> have been performed using various ML techniques</p>	<p>Classification</p>	<ul style="list-style-type: none"> • Naïve Bayes • Decision trees • Support vector machine • Artificial neural network 	<p>Chew and Robinson (2012), Samoil (2015)</p>
<p><i>Process 3: Preparation of management accounts</i> <i>Task 3.3 Account allocation</i> <i>Account allocation</i> may be performed by incorporating ML, which learns to predict the account allocation based on probability and can recommend which accounts to post to</p>	<p>Classification and clustering</p>	<p><i>Classification:</i> Naive Bayes</p> <p><i>Clustering:</i> K-means clustering Random forests</p>	<p>Bengtsson and Jansson (2015), Brady <i>et al.</i> (2017), SMACC (2017), Takaki and Ericson (2018)</p>

(continued)

Description of the learning problem	Solutions to learning problem	ML techniques	Source
<p><i>Task 3.7 Report generation</i></p> <p><i>Error detection in financial data and fraud detection</i> can be performed by incorporating ML to identify irregularities in data sets</p>	<p>Solutions to learning problem</p> <p>Classification; outlier detection and clustering</p>	<p><i>Classification:</i></p> <ul style="list-style-type: none"> Bayesian belief network and a decision table Naive Bayes hybrid model <p><i>Outlier detection:</i></p> <ul style="list-style-type: none"> Association rules <p><i>Clustering:</i></p> <ul style="list-style-type: none"> K-means clustering Self-organising maps 	<p>Ahmed <i>et al.</i> (2016), Alpar and Winkelsträter (2014), Hajek and Henriques (2017), Kokina and Davenport (2017)</p>
<p><i>Task 3.8 Report descriptions</i></p> <p><i>Report descriptions</i> may incorporate ML techniques in natural language generation technologies to enable a reasoning process to be applied to the reported data to produce required explanations in natural language</p>	<p>Prediction</p>	<p>Conditional random fields</p>	<p>Gardent and Perez-Beltrachini (2017), Lafferty <i>et al.</i> (2001), Yseop (2017)</p>

Source: Compiled by author from multiple sources as indicated above

Table 3.

Table 4.
Description of ML
limitations

Limitation	Description of limitation	Source	Qualitative accounting objective
Poor interpretability	The limitation relates to the fact that users cannot understand how information is generated by the ML technology owing to the complexity of the ML model	Ayoadele (2010a), Sainani (2014), The Royal Society (2017)	Verifiability and Understandability
Overfitting	The risk is that input features with little modelling benefit are included in the training data. These features may increase the sensitivity of the technology to changes in the inputs, even though they could be excluded with no disadvantages. In such instances, the ML model may be too closely linked to the training data used to train it and unable to classify other data sets appropriately. This increases the risk of misleading representations	Hawkins (2004), Sculley <i>et al.</i> (2015), Witten <i>et al.</i> (2016)	Relevance
It takes a long time to train	The risk of increased training or learning times for ML models as the size and complexity of the data sets increase	Ghanem (2012)	Timeliness
Complex, which makes it slow	The risk of increased processing time due to the complexity of the ML model. In the case of a classification technique, for example, the model will be slow to classify data	Kotsiantis (2007), Witten <i>et al.</i> (2016)	Timeliness
Training rate may be slow depending on available labelled data	In this instance, the training rate is impacted by the available labelled data to train the ML model. The ML model is trained using a labelled data set consisting of examples of input data and labels indicating predicted targets or output data. This type of data is not as prevalent as unstructured data	Ayoadele (2010b), Castle (2018), Lausson and Segerås (2016), Marsland (2009), SMIACC (2017), Zheng <i>et al.</i> (2017)	Timeliness
Requires independent variables	An ML technique such as Naïve Bayes requires independent variables in the data, which implies that the values of the different features of each variable do not influence one another. However, as each variable has a high number of features, it is unlikely that there are no dependencies among them. This may result in incorrect processing and outputs of the ML model	Marsland (2009), Samoil (2015)	Faithful representation

(continued)

Limitation	Description of limitation	Source	Qualitative accounting objective
Training set sensitive	If a feature has a category which was not observed in the training data set, then a zero probability will be assigned to that category, thus resulting in the ML model not being able to make a prediction known as zero frequency	Witten <i>et al.</i> (2016), Samoil (2015), Larsson and Segerås (2016)	Materiality and faithful representation
Computing intensive	The risk of costs exceeding the financial benefits to the business as the ML model requires advanced data integration tools and infrastructure, which may present significant costs to the business to acquire	Gillion (2017), Sapp (2017)	Cost-saving
Excessive output	In the case of association rules ML models, the number of rules discovered may be excessive, which may impact the relevance of the output	Kaur (2014)	Relevance
Requires lots of time	ML models may take a lot of time to produce outputs, and this may be due to how many times the algorithm needs to run to achieve an accurate result	Witten <i>et al.</i> (2016), Kaur (2014), Ayodele (2010b)	Timeliness
Requires adequate data	The risk is that the ML model is inaccurate owing to insufficient data for training	Burrell (2016)	Materiality and faithful representation
Trade-off between accuracy, which requires memory and overfitting	A limitation of some ML models is that training them using large feature data sets results in more accurate predictions but requires more memory to store and has an increased risk of overfitting	Witten <i>et al.</i> (2016), Sutton and McCallum (2011)	Relevance and faithful representation

Notes: We have inserted the Table here for ease of review. We have kept to the standard *JFRA* convention in the “clean copy” version of the manuscript

Table 4.

Table 5.
Limitations of ML
techniques mapped
to objectives and
tasks

ML techniques	Limitations	Qualitative accounting objective	Tasks in the accounting process	Source
Transfer learning decision forests and random forests	<ul style="list-style-type: none"> Poor interpretability Overfitting 	<p>Verifiability and understandability</p> <p>Relevance</p>	<ul style="list-style-type: none"> Adaptability of OCR Account allocation 	Dataiku (2017)
Support vector machine	<ul style="list-style-type: none"> Poor interpretability 	<p>Verifiability and understandability</p>	<ul style="list-style-type: none"> Image classification Validation of document information Removing duplicate entries and linking documents Matching records or record-linkage 	Kotsiantis (2007), Karamzadeh <i>et al.</i> (2014), Witten <i>et al.</i> (2016)
Convolutional neural networks	<ul style="list-style-type: none"> Poor interpretability Takes a long time to train 	<p>Verifiability and understandability</p> <p>Timeliness</p>	<ul style="list-style-type: none"> Deep document feature extraction Irregular document layout classification using NLP 	Dataiku (2017), Tarawneh <i>et al.</i> (2019)
<i>k</i> -Nearest neighbour	<ul style="list-style-type: none"> Poor interpretability Complex, which makes it slow 	<p>Verifiability and understandability</p> <p>Timeliness</p>	<ul style="list-style-type: none"> Image classification Text classification 	Kotsiantis (2007), Witten <i>et al.</i> (2016)
Semi-supervised clustering	<ul style="list-style-type: none"> Training rate may be slow depending on available labelled data 	<p>Timeliness</p>	<ul style="list-style-type: none"> Text classification 	Zheng, Zhou, Deng, <i>et al.</i> (2017)
Naïve Bayes	<ul style="list-style-type: none"> Requires independent variables Training set sensitive 	<p>Faithful representation</p> <p>Materiality and faithful representation</p>	<ul style="list-style-type: none"> Validation of document information Removing duplicate entries and linking documents Matching records or record-linkage Account allocation 	Samoil (2015) Witten <i>et al.</i> (2016), Samoil (2015), Larsson and Segerås (2016)

(continued)

ML techniques	Limitations	Qualitative accounting objective	Tasks in the accounting process	Source
Artificial neural network	<ul style="list-style-type: none"> Poor interpretability Computing intensive 	<ul style="list-style-type: none"> Verifiability and understandability Cost-saving 	<ul style="list-style-type: none"> Error detection in financial data and fraud detection Removing duplicate entries and linking documents Matching records or record-linkage 	Koisiantis (2007), Dataiku (2017), SMACC (2017)
Bayesian Belief network	<ul style="list-style-type: none"> Computing intensive Training set sensitive 	<ul style="list-style-type: none"> Cost-saving Materiality and faithful representation 	<ul style="list-style-type: none"> Error detection in financial data and fraud detection 	Heckerman (2008), Niedermayer (2008)
Association rules	<ul style="list-style-type: none"> Excessive output Requires lots of time 	<ul style="list-style-type: none"> Relevance Timeliness 	<ul style="list-style-type: none"> Error detection in financial data and fraud detection 	Witten <i>et al.</i> (2016), Kaur (2014)
K-means clustering	<ul style="list-style-type: none"> Requires lots of time 	<ul style="list-style-type: none"> Timeliness 	<ul style="list-style-type: none"> Account allocation Error detection in financial data and fraud detection 	Ayodele (2010b), Witten <i>et al.</i> (2016)
Self-organising maps	<ul style="list-style-type: none"> Computing intensive Requires adequate data 	<ul style="list-style-type: none"> Cost-saving Materiality and faithful representation 	<ul style="list-style-type: none"> Error detection in financial data and fraud detection 	Ayodele (2010b), SMACC (2017)
Conditional random fields	<ul style="list-style-type: none"> Trade-off between accuracy, which requires memory and overfitting 	<ul style="list-style-type: none"> Relevance and faithful representation 	<ul style="list-style-type: none"> Report descriptions 	Witten <i>et al.</i> (2016), Sutton and McCallum (2007)

Source: Compiled by author from multiple sources as indicated above

Table 5.

The primary limitations identified from the research for the respective ML techniques suitable for use in the selected accounting processes have been summarised in [Table 5](#) and linked to the appropriate accounting objectives.

As noted in [Table 5](#), some ML techniques limit interpretability, as the knowledge that the ML technique uses or discovers to perform a particular task may not always be available to the user ([The Royal Society, 2017](#)). This would mean that the user does not know how data input A resulted in information output B. This impacts the accounting objectives of verifiability and understandability.

The above could be a problem when using, for example, an artificial neural network ML technique to perform matching during a reconciliation accounting task, especially if it is unclear why the application matched one transaction with another.

Another limitation of some ML techniques is that they are quite computing intensive. This aspect of ML and AI has been identified as one of the challenges for the metaverse, as it impacts access to users on mobile devices, for example ([Lee et al., 2021](#)). This limitation will need to be considered in terms of whether the costs of incorporating this technology in the accounting process may exceed the financial benefits to the business as these ML techniques require advanced data integration tools and infrastructure ([Gillion, 2017](#); [Sapp, 2017](#)). [Blau et al. \(2022\)](#) advised users wanting to incorporate the metaverse into their strategy to consider such costs from a broader digital transformation perspective, not just the financial benefit of investing in ML technology.

The limitation of adequate data being required for the training of algorithms may not be applicable in the metaverse, as one of the characteristics of the metaverse would be the availability of immeasurable amounts of structured data ([Lee et al., 2021](#)).

Conclusion

This paper aimed to identify the limitations of those ML techniques that can perform specific accounting tasks within the accounting process. We achieved our aim by identifying the accounting tasks that ML could perform, which techniques would be able to perform which functions in the accounting tasks and then identifying the limitations associated with specific ML techniques. Finally, these limitations were linked to qualitative accounting objectives, which may be impacted.

We find that there were limitations to the ML techniques, which may impact the achievement of the qualitative accounting objectives when using ML in the accounting process. Some ML limitations identified in this research are barriers to entry into the metaverse, such as the extent of the computation abilities required to apply ML. In contrast, other limitations, such as adequate data, may be addressed by immersion in the metaverse.

Our research highlights the need for accounting users to understand where ML technology can be used in the accounting process and to be aware of the limitations of ML techniques that may impact the ability to achieve qualitative accounting objectives. As tasks are transformed from the physical to the digital world, we, as accounting users, can lead in enabling a governed transition to the metaverse.

Limitations and future research

The study does not intend to cover all the areas of accounting in which ML intervention is possible; it only addressed the three accounting processes: the translation of manual and electronic documents into accounting software information, the reconciliation of financial information and the preparation of management accounts. Areas for future research could consider the use of ML in accounting estimates, such as the estimate of expected credit losses of loans in the financial services industry.

Only those tasks for which a suitable ML technique can be found during this research were addressed. Furthermore, the study only considered ML techniques appropriate for addressing the identified accounting tasks in the accounting process and, therefore, does not intend to present an exhaustive list of ML techniques.

Our study does not explain how each ML technique functions. The limitations identified for the ML techniques are those unique to the technology and not those about the environment in which ML operates, such as a database or accounting software limitations or limitations about supporting technologies such as cloud platforms. Hence, these limitations are not addressed in this paper. Furthermore, considerable research has been performed in the areas of ML applied to auditing and the detection of fraud using such technologies. Therefore, these areas were not considered for this research.

Future research could look at how these limitations can be addressed to ensure that the technology is adequately governed. The question of whether standards (such as specifications regarding training data, ML model, performance and updating of models) need to be set to regulate the ML tools used by accountants (Cho *et al.*, 2020) has also been asked, considering the limitations and risks faced when using ML technology and the need for technology governance. Considering the risks, opportunities for further research may also include a better understanding of the risks when implementing ML technology in an accounting system. Research into the benefits of ML can moderate future research into the limitations, as explained above.

Finally, in this study, we consider ML as one of the technologies enabling rapid digital transformation to the metaverse, explicitly focusing on the limitations of ML techniques in an accounting context. Future research could explore some of the challenges and opportunities that the rapid pace of digital transformation presents for technology governance in the accounting field.

References

- ABBYY Technologies (2017), "ABBYY FlexiCapture", available at: www.abbyy.com/en-apac/flexicapture/features/ (accessed 2 August 2017).
- Ahmed, M., Mahmood, A.N. and Islam, M.R. (2016), "A survey of anomaly detection techniques in financial domain", *Future Generation Computer Systems*, Vol. 55, pp. 278-288, doi: [10.1016/j.future.2015.01.001](https://doi.org/10.1016/j.future.2015.01.001).
- Algorithmia (2020), "2020 State of enterprise machine learning", available at: https://info.algorithmia.com/hubfs/2019/Whitepapers/The-State-of-Enterprise-ML-2020/Algorithmia_2020_State_of_Enterprise_ML.pdf%0Ahttps://algorithmia.com/state-of-ml (accessed 10 June 2022).
- Alpar, P. and Winkelsträter, S. (2014), "Assessment of data quality in accounting data with association rules", *Expert Systems with Applications*, Vol. 41 No. 5, pp. 2259-2268, doi: [10.1016/j.eswa.2013.09.024](https://doi.org/10.1016/j.eswa.2013.09.024).
- Alreemy, Z., Chang, V., Walters, R. and Wills, G. (2016), "Critical success factors (CSFs) for information technology governance (ITG)", *International Journal of Information Management*, Vol. 36 No. 6, pp. 907-916, doi: [10.1016/j.ijinfomgt.2016.05.017](https://doi.org/10.1016/j.ijinfomgt.2016.05.017).
- Amani, F.A. and Fadlalla, A.M. (2017), "Data mining applications in accounting: a review of the literature and organizing framework", *International Journal of Accounting Information Systems*, Vol. 24, pp. 32-58, doi: [10.1016/j.accinf.2016.12.004](https://doi.org/10.1016/j.accinf.2016.12.004).
- Amtrup, J.W., Thompson, S.M., Kilby, S. and Macciola, A. (2015), *Patent No. 9058580 B1*, United States.
- Ayodele, T.O. (2010a), "Introduction to machine learning", in Zhang, Y. (Ed.), *New Advances in Machine Learning*, InTech, Croatia, pp. 1-8, available from: www.intechopen.com/books/new-advances-in-machine-learning/introduction-to-machine-learning

- Ayodele, T.O. (2010b), "Types of machine learning algorithms", in Zhang, Y. (Ed), *New Advances in Machine Learning*, InTech, Croatia, pp. 19-48, available at: www.intechopen.com/books/new-advances-in-machine-learning/types-of-machine-learning-algorithms
- Barocas, S. and Selbst, A.D. (2016), "Big data's disparate impact", *California Law Review*, Vol. 104 No. 3, pp. 671-732, available at: <https://lawcat.berkeley.edu/record/1127463>
- Bavaresco, R.S., Nesi, L.C., Barbosa, J.L.V., Antunes, R.S., da Rosa Righi, R., da Costa, C.A., Vanzin, M., Dornelles, D., Gatti, C., Ferreira, M. and Silva, E. (2023), "Machine learning-based automation of accounting services: an exploratory case study", *International Journal of Accounting Information Systems*, Vol. 49, p. 100618.
- Bengtsson, H. and Jansson, J. (2015), *Using Classification Algorithms for Smart Suggestions in Accounting Systems*, Chalmers University of Technology, Gothenburg.
- Blau, A., Enobakhare, A.J., Shiller, A., Lubetsky, L. and Walker, M.W. (2022), "A whole new world? Exploring the metaverse and what it could mean for you", available at: www2.deloitte.com/content/dam/Deloitte/us/Documents/technology/us-ai-institute-what-is-the-metaverse-new.pdf (accessed 29 May 2023).
- Brady, E.S., Leider, J.P., Resnick, B.A., Natalia Alfonso, Y. and Bishai, D. (2017), "Machine-learning algorithms to code public health spending accounts", *Public Health Reports*, Vol. 132 No. 3, pp. 350-356, doi: [10.1177/0033354917700356](https://doi.org/10.1177/0033354917700356).
- Bryant, A. (2002), "Re-grounding grounded theory", *Journal of Information Technology Theory and Application*, Vol. 4 No. 1, pp. 25-42.
- Bucheli, H. and Thompson, W. (2014), "Statistics and machine learning at scale", available at: https://media.bitpipe.com/io_12x/io_123306/item_1144704/StatisticsandMachineLearningatScale-discovery.pdf (accessed 27 May 2018).
- Burrell, J. (2016), "How the machine 'thinks': understanding opacity in machine learning algorithms", *Big Data and Society*, Vol. 3 No. 1, pp. 1-12.
- Castle, N. (2018), "What is Semi-Supervised learning?", available at: www.datascience.com/blog/what-is-semi-supervised-learning (accessed 15 February 2018).
- Caton, S. and Haas, C. (2023), "Fairness in machine learning: a survey", *ACM Computing Surveys*, Vol. 56 No. 7, pp. 1-38, doi: [10.1145/3616865](https://doi.org/10.1145/3616865).
- Chen, L., Wang, S., Fan, W., Sun, J. and Satoshi, N. (2015), "Deep learning based language and orientation recognition in document analysis", 2015 13th International Conference on Document Analysis and Recognition, *Tunis, Tunisia*, IEEE. pp. 436-440, [10.1109/ICDAR.2015.7333799](https://doi.org/10.1109/ICDAR.2015.7333799).
- Chew, P.A. and Robinson, D.G. (2012), "Automated account reconciliation using probabilistic and statistical techniques", *International Journal of Accounting and Information Management*, Vol. 20 No. 4, pp. 322-334.
- Cho, S., Vasarhelyi, M.A., Sun, T. and Zhang, C. (2020), "Learning from machine learning in accounting and assurance", *Journal of Emerging Technologies in Accounting*, Vol. 17 No. 1, pp. 1-10, doi: [10.2308/jeta-10718](https://doi.org/10.2308/jeta-10718).
- Dataiku (2017), "Machine learning basics: an illustrated guide for Non-Technical readers", available at: <https://pages.dataiku.com/machine-learning-basics-thank-you?submissionGuid=80d21f82-ac46-45d0-969a-cd9914d06af9> (accessed 6 February 2018).
- De Leone, R. and Minnetti, V. (2015), "Electre Tri-Machine learning approach to the record linkage problem".
- Deloitte (2020), "What North America's top finance executives are thinking - and doing", available at: www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/CFO_Signals_3Q20_Report.pdf (accessed 10 June 2022).
- Deming, Z. (2024), "The modern approach to closing the books - An introduction to continuous accounting", available at: www.blackline.com/assets/docs/uploads/continuous-accounting-ebook.pdf (accessed 21 November 2017).

- Desai, D., Jain, A., Naik, D., Panchal, N. and Sawant, D. (2021), "Invoice processing using RPA and AI", International Conference on Smart Data Intelligence, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3852575
- Du, J. (2017), "Automatic text classification algorithm based on gauss improved convolutional neural network", *Journal of Computational Science*, Vol. 21, pp. 195-200, doi: [10.1016/j.jocs.2017.06.010](https://doi.org/10.1016/j.jocs.2017.06.010).
- Everest Group (2014), "Service delivery automation (SDA) market in 2014: moving business process services beyond labor arbitrage", available at: www.everestgrp.com/wp-content/uploads/2014/10/Service-Delivery-Automation-Market-in-2014-Everest-Group-Report.pdf (accessed 15 November 2017).
- Fallatah, R. (2021), "Machine learning in production: a literature review", *Academy of Accounting and Financial Studies Journal*, Vol. 25 No. 7, pp. 1-14.
- Fernandez, W.D. and Lehmann, H. (2011), "Case studies and grounded theory method in information systems research: issues and use", *Journal of Information Technology Case and Application Research*, Vol. 13 No. 1, pp. 4-15, doi: [10.1080/15228053.2011.10856199](https://doi.org/10.1080/15228053.2011.10856199).
- Gardent, C. and Perez-Beltrachini, L. (2017), "A statistical, Grammar-Based approach to microplanning", *Computational Linguistics*, Vol. 43 No. 1, pp. 1-30, doi: [10.1162/COLI](https://doi.org/10.1162/COLI).
- Ghanem, K. (2012), "A simple process to speed up machine learning methods: application to hidden Markov models", *Computer Science and Information Technology (CS & IT) – Computer Science Conference Proceedings (CSCP)*, Vol. 2 No. 5, pp. 161-171, available at: <https://airccj.org/CSCP/vol2/csit2514.pdf>
- Gillion, K. (2017), "Artificial intelligence and the future of accountancy", doi: [10.1145/2063176.2063177](https://doi.org/10.1145/2063176.2063177).
- Goosen, R. and Rudman, R.J. (2013), "An integrated framework to implement IT governance principles at a strategic and operational level for medium to large sized South African businesses", *International Business and Economics Research Journal (IBER)*, Vol. 12 No. 7, pp. 835-854.
- Goussies, N.A., Ubalde, S., Fernandez, F.G. and Mejail, M.E. (2014), "Optical character recognition using transfer learning decision forests", Paris, France, IEEE, available at: <http://ieeexplore.ieee.org.ez.sun.ac.za/document/7025875/> (accessed 27 August 2017).
- Hadzic, F., Dillon, T.S. and Tan, H. (2007), "Outlier detection strategy using self-organising maps", in Zhu, X. and Davidson, I. (Eds). *Knowledge Discovery and Data Mining: Challenges and Realities*, IGI Global, Hershey, pp. 224-243.
- Hajek, P. and Henriques, R. (2017), "Mining corporate annual reports for intelligent detection of financial statement fraud: a comparative study of machine learning methods", *Knowledge-Based Systems*, Vol. 128 No. 2017, pp. 139-152, doi: [10.1016/j.knosys.2017.05.001](https://doi.org/10.1016/j.knosys.2017.05.001).
- Hawkins, D.M. (2004), "The problem of overfitting", *Journal of Chemical Information and Computer Sciences*, Vol. 44 No. 1, pp. 1-12.
- Heckerman, D. (2008), "Innovations in Bayesian networks", in Holmes, D.E. and Jain, L.C. (Eds.), *Innovations in Bayesian Networks*, Springer-Verlag, Berlin Heidelberg, pp. 33-82, doi: [10.1007/978-3-540-85066-3](https://doi.org/10.1007/978-3-540-85066-3).
- Institute of Directors of Southern Africa (IODSA) (2016), "King IV report on corporate governance for South Africa 2016", available at: https://c.yimcdn.com/sites/www.iodsa.co.za/resource/resmgr/king_iv/King_IV_Report/IoDSA_King_IV_Report_-_WebVe.pdf (accessed 13 May 2018).
- International Accounting Standards Board. (2022), "Conceptual framework for financial reporting", IFRS Foundation.
- Jain, A., Jin, R. and Chitta, R. (2015), "Semi-supervised clustering", in Hennig, C., Meila, M., Murtagh, F. and Rocci, R. (Eds), *Handbook of Cluster Analysis*, Chapman and Hall/CRC, New York, pp. 443-469, doi: [10.1201/b19706](https://doi.org/10.1201/b19706).
- Karamzadeh, S., Abdullah, S.M., Halimi, M., Shayan, J. and Rajabi, M.J. (2014), "Advantage and drawback of support vector machine functionality", 1st International Conference on Computer, Communications, and Control Technology, Proceedings, pp. 63-65, [10.1109/I4CT.2014.6914146](https://doi.org/10.1109/I4CT.2014.6914146).

- Kaur, G. (2014), "Association rule mining: a survey. Vol. 5", available at: <http://sci2s.ugr.es/keel/pdf/specific/report/zhao03ars.pdf> (accessed 04 June 2018).
- Khan, A. (2019), "Comparison of machine learning approaches for classification of invoices", Unpublished master's thesis. Tampere, Tampere University, doi: [10.7717/peerj.10549](https://doi.org/10.7717/peerj.10549).
- Kleinberg, J., Mullainathan, S. and Raghavan, M. (2016), "Inherent trade-offs in the fair determination of risk scores", 8th Innovations in Theoretical Computer Science Conference (ITCS 2017). Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl – Leibniz-Zentrum für Informatik (2017), Vol. 67, pp. 43:1-43:23, doi: [10.4230/LIPIcs.ITCS.2017.43](https://doi.org/10.4230/LIPIcs.ITCS.2017.43).
- Kohonen, T. (1990), "The self-organizing map", *Proceedings of the IEEE*, Vol. 78 No. 9, pp. 1464-1480, doi: [10.1109/5.58325](https://doi.org/10.1109/5.58325).
- Kokina, J. and Davenport, T.H. (2017), "The emergence of artificial intelligence: how automation is changing auditing", *Journal of Emerging Technologies in Accounting*, Vol. 14 No. 1, pp. 115-122, doi: [10.2308/jeta-51730](https://doi.org/10.2308/jeta-51730).
- Kommunuri, J. (2022), "Artificial intelligence and the changing landscape of accounting: a viewpoint", *Pacific Accounting Review*, Vol. 34 No. 4, pp. 585-594.
- Kotsiantis, S.B. (2007), "Supervised machine learning: a review of classification techniques", *Informatica*, Vol. 31, pp. 249-268, doi: [10.1115/1.1559160](https://doi.org/10.1115/1.1559160).
- Lafferty, J., McCallum, A. and Pereira, F. (2001), "Conditional random fields: probabilistic models for segmenting and labeling sequence data", Proceedings of the 18th International Conference on Machine Learning: (ICML), Vol. 1, pp. 282-289.
- Larsson, A. and Segerås, T. (2016), "Automated invoice handling with machine learning and OCR", available at: www.diva-portal.org/smash/get/diva2:934351/FULLTEXT01.pdf (accessed 07 June 2017).
- Lee, L.-H., Braud, T., Zhou, P., Wang, L., Xu, D., Lin, Z., Kumar, A., Bermejo, C. and Hui, P. (2021), "All one needs to know about metaverse: a complete survey on technological singularity, virtual ecosystem, and research agenda. (October, 6)", available at: <http://arxiv.org/abs/2110.05352>
- Mardini, G.H. and Alkurdi, A. (2021), "Artificial intelligence literature in accounting: a panel systematic approach", *The Fourth Industrial Revolution: Implementation of Artificial Intelligence for Growing Business Success*, Springer International Publishing, Cham, pp. 311-323.
- Marsland, S. (2009), *Machine Learning an Algorithmic Perspective*, CRC Press, Boca Raton.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021), "A survey on bias and fairness in machine learning", *ACM Computing Surveys*, Vol. 54 No. 6, pp. 1-35, doi: [10.1145/3457607](https://doi.org/10.1145/3457607).
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D. and Gebru, T. (2019), "Model cards for model reporting", FAT* '19: Conference on Fairness, Accountability, and Transparency, January 29–31, 2019, Atlanta, GA, USA, ACM, New York, NY, ten pages, doi: [10.1145/3287560.3287596](https://doi.org/10.1145/3287560.3287596).
- Narasimha Murty, M. and Susheela Devi, V. (2011), *Pattern Recognition: An Algorithmic Approach*, in Mackie, I. (Ed.). Springer, London.
- Niedermayer, D. (2008), "An introduction to Bayesian networks and their contemporary applications", *Studies in Computational Intelligence*, Vol. 156, pp. 117-129, doi: [10.1007/978-3-540-85066-3_5](https://doi.org/10.1007/978-3-540-85066-3_5).
- Oquab, M., Bottou, L., Laptev, I. and Sivic, J. (2014), "Learning and transferring mid-level image representations using convolutional neural networks", *2014 IEEE Conference Computer Vision and Pattern Recognition, Columbus, OH, USA*, IEEE, pp. 1717-1724, doi: [10.1109/CVPR.2014.222](https://doi.org/10.1109/CVPR.2014.222).
- Pandey, D. and Gilmour, P. (2023), "Accounting meets metaverse: navigating the intersection between the real and virtual worlds", *Journal of Financial Reporting and Accounting*, doi: [10.1108/JFRA-03-2023-0157](https://doi.org/10.1108/JFRA-03-2023-0157).
- Petkov, R. (2020), "Artificial intelligence and the accounting function. A revisit and a new perspective for developing framework", *Journal of Emerging Technologies in Accounting*, Vol. 17 No. 1, pp. 99-105. doi: [10.2308/jeta-52648](https://doi.org/10.2308/jeta-52648)

- PWC (2019), "A practical guide to responsible artificial intelligence", available at: www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/responsible-ai-practical-guide.pdf (accessed 21 April 2021).
- Sainani, K.L. (2014), "Explanatory versus predictive modeling", *PM&R*, Vol. 6 No. 9, pp. 841-844, doi: [10.1016/j.pmrj.2014.08.941](https://doi.org/10.1016/j.pmrj.2014.08.941).
- Saitta, L. and Neri, F. (1998), "Learning in the 'real world'", *Machine Learning*, Vol. 30 Nos 2/3, pp. 133-163, doi: [10.1023/A:1007448122119](https://doi.org/10.1023/A:1007448122119)
- Samoil, L.A. (2015), *Multiple Entity Reconciliation*, KTH Royal Institute of Technology, Stockholm.
- Sapp, C.E. (2017), "Preparing and architecting for machine learning", doi: G00317328.
- Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.F. and Dennison, D. (2015), "Hidden technical debt in machine learning systems", available at: <http://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf> (accessed 20 May 2018).
- SMACC (2017), "Are SMEs able to keep an eye on their corporate finances in the same way as large corporates do?", available at: www.smacc.io/wp-content/uploads/2017/04/finance-with-artificial-intelligence.pdf (accessed 05 February 2018).
- Someren, M.V. and Urbancic, T. (2006), "Applications of machine learning: matching problems to tasks and methods", *The Knowledge Engineering Review*, Vol. 20 No. 4, pp. 363-402, doi: [10.1017/S0269888906000762](https://doi.org/10.1017/S0269888906000762).
- Sorio, E. (2013), *Machine Learning Techniques for Document Processing and Web Security*, University of Trieste, Trieste.
- Sorio, E., Bartoli, A., Davanzo, G. and Medvet, E. (2010), "Open world classification of printed invoices", Proceedings of the 10th ACM symposium on Document engineering, *Manchester, United Kingdom*, ACM, pp. 187-190, [10.1145/1860559.1860599](https://doi.org/10.1145/1860559.1860599).
- Suresh, H. and Gutttag, J.V. (2019), "A framework for understanding unintended consequences of machine learning", available at: <https://arxiv.org/abs/1901.10002>
- Sutton, C. and McCallum, A. (2007), "An introduction to conditional random fields for relational learning", in Getoor, L. and Taskar, B. (Eds). *Introduction to Statistical Relational Learning*, The MIT Press, Cambridge, MA, pp. 93-128, doi: [10.1677/JME-08-0087](https://doi.org/10.1677/JME-08-0087).
- Sutton, C. and McCallum, A. (2011), "An introduction to conditional random fields", *Foundations and Trends® in Machine Learning*, Vol. 4 No. 4, pp. 267-373, doi: [10.1561/22000000013](https://doi.org/10.1561/22000000013).
- Sutton, S.G., Holt, M. and Arnold, V. (2016), "The reports of my death are greatly exaggerated": artificial intelligence research in accounting", *International Journal of Accounting Information Systems*, Vol. 22, pp. 60-73.
- Sutton, S.G., Reinking, J. and Arnold, V. (2011), "On the use of grounded theory as a basis for research on strategic and emerging technologies in accounting", *Journal of Emerging Technologies in Accounting*, Vol. 8 No. 1, pp. 45-63, doi: [10.2308/jeta-10207](https://doi.org/10.2308/jeta-10207).
- Takaki, J. and Ericson, G. (2018), "Assign data to clusters", Available, available at: <https://cloud.google.com/blog/big-data/2018/01/problem-solving-with-ml-automatic-document-classification> (accessed 07 February 2018).
- Tarawneh, A.S., Hassanat, A.B., Chetverikov, D., Lendak, I. and Verma, C. (2019), "Invoice classification using deep features and machine learning techniques", 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT 2019 – Proceedings, IEEE, pp. 855-859, doi: [10.1109/JEEIT.2019.8717504](https://doi.org/10.1109/JEEIT.2019.8717504).
- The Royal Society (2017), *Machine Learning: The Power and Promise of Computers That Learn by Example*, The Royal Society, London, doi: [10.1126/scitranslmed.3002564](https://doi.org/10.1126/scitranslmed.3002564).
- Thomassey, S. and Fiordaliso, A. (2006), "A hybrid sales forecasting system based on clustering and decision trees", *Decision Support Systems*, Vol. 42 No. 1, pp. 408-421, doi: [10.1016/j.dss.2005.01.008](https://doi.org/10.1016/j.dss.2005.01.008).

- USC Libraries. (2024), "Organizing your social sciences research paper: types of research designs", available at: <http://libguides.usc.edu/writingguide/researchdesigns> (accessed 18 January 2018).
- Valavanis, K.P., Kokkinaki, A.I. and Tzafestas, S.G. (1994), "Knowledge-based (expert) systems in engineering applications: a survey", *Journal of Intelligent and Robotic Systems*, Vol. 10 No. 2, pp. 113-145.
- Vial, G. (2019), "Understanding digital transformation: a review and a research agenda", *The Journal of Strategic Information Systems*, Vol. 28 No. 2, pp. 118-144, doi: [10.1016/j.jsis.2019.01.003](https://doi.org/10.1016/j.jsis.2019.01.003).
- Wilkin, C.L. and Chenhall, R.H. (2020), "Information technology governance: reflections on the past and future directions", *Journal of Information Systems*, Vol. 34 No. 2, pp. 257-292, doi: [10.2308/isy-52632](https://doi.org/10.2308/isy-52632).
- Winkler, W.E. (2014), "Matching and record linkage", *WIREs Computational Statistics*, Vol. 6 No. 5, pp. 313-325.
- Witten, I.H., Frank, E., Hall, M.A. and Pal, C.J. (2016), *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed., Morgan Kaufman, Cambridge.
- Wolfswinkel, J.F., Furttmueller, E. and Wilderom, C.P.M. (2013), "Using grounded theory as a method for rigorously reviewing literature", *European Journal of Information Systems*, Vol. 22 No. 1, pp. 45-55, doi: [10.1057/ejis.2011.51](https://doi.org/10.1057/ejis.2011.51).
- World Economic Forum. (2021), "The global risks report 2021: 16th edition", available at: www3.weforum.org/docs/WEF_The_Global_Risks_Report_2021.pdf (accessed 22 April 2021).
- Yseop (2017), "Integrating natural language generation software into finance workflows", available at: <http://compose.yseop.com/downloads/integrating-natural-language-generation-finance-workflows/> (accessed 10 September 2018).
- Zhang, W., Tang, X. and Yoshida, T. (2015), "TESC: an approach to text classification using semi-supervised clustering", *Knowledge-Based Systems*, Vol. 75, pp. 152-160, doi: [10.1016/j.knosys.2014.11.028](https://doi.org/10.1016/j.knosys.2014.11.028).
- Zheng, J., Zhou, Y., Deng, T. and Yang, X. (2017), "A self-trained semi supervised fuzzy clustering based on label propagation with variable weights", *Proceedings of the 29th Chinese Control and Decision Conference, CCDC 2017*, pp. 7447-7452, doi: [10.1109/CCDC.2017.7978533](https://doi.org/10.1109/CCDC.2017.7978533).

Further reading

- BlackLine (2014), "What are account reconciliations", available at: www.blackline.com/blog/account-reconciliation/ (accessed 11 June 2017).
- DBASS Chartered Accountants (2014), "How to produce valuable management accounts using sage line 50", available at: www.dbass.ie/advice-resources/advice/how-to-produce-valuable-management-accounts-using-sage-line-50 (accessed 20 August 2017).
- Ming, D., Liu, J. and Tian, J. (2003), "Research on Chinese financial invoice recognition technology", *Pattern Recognition Letters*, Vol. 24 Nos 1/3, pp. 489-497, doi: [10.1016/S0167-8655\(02\)00271-4](https://doi.org/10.1016/S0167-8655(02)00271-4).

Corresponding author

Christiaan Lamprecht can be contacted at: clam@sun.ac.za

ML solution	Algorithm	Description	Source
<i>Classification [Supervised]</i>	<i>Logic-based</i> Decision trees	The decision tree consists of nodes, each containing a question related to a particular feature. The algorithm starts at the root node, determines which features are present for that root node question and then, depending on the answer, moves on to the next node. The classification consists of a number of decisions which occur at each node, ending at the leaf node An iteration of the normal decision tree algorithm by pruning it to reduce the number of nodes without losing the ability to classify the instance This model uses random forests (see Clustering, unsupervised) where the knowledge produced can subsequently be applied or transferred to a given target task. This generates a classifier that can be used to exploit the knowledge from other tasks to improve the ability of the classifier to perform a target task	Thomassey and Fiordaliso (2006:410), Koitsiantis (2007:251), Marsland (2009:133), Narasimha Murty and Susheela Devi (2011:127)
	C4.5 decision trees		Thomassey and Fiordaliso (2006:410), Marsland (2009:143) Goussies <i>et al.</i> (2014:4312
	Random forests (transfer learning decision forests)		SMACC (2017:9)
	<i>Perceptron-based</i> Neural network (dual-use algorithm)	A combination of mathematically generated neurons, which operate similarly to the human brain. These neurons are each assigned a weight based on what the artificial neural network learns, collectively forming part of a mathematical function. A network consisting of an input and an output layer can classify linearly separable classes – this is known as a feed-forward network	Larsson and Segerås (2016:12)
	<i>Statistical</i> Naïve Bayes	The Naïve Bayes algorithm is a probabilistic model which determines the probability of different classes or outcomes based on previously encountered examples These examples are identified in the training data The network is modelled on the Bayesian theorem. Using a graphical model, it considers the probabilistic dependencies among features and plots these probabilistic relationships	Heckerman (2008:33), Narasimha Murty and Susheela Devi (2011:97), Witten <i>et al.</i> (2016:340)
	Bayesian belief networks		(<i>continued</i>)

Table A1.
Description of ML techniques

Table A1.

ML solution	Algorithm	Description	Source
<i>Prediction [Supervised]</i>	K-nearest neighbour	The algorithm classifies instances or patterns according to the nearest known neighbour class by finding similarities in the instance being classified to patterns or features in the training set	Narasimha Murty and Susheela Devi (2011:48)
	Support vector machines (dual-use algorithm) Conditional random fields	A binary classifier that aims to separate data into two classes based on the case features An algorithm can predict outputs by combining discriminative classification with graphical modelling	Ayodele (2010b:25) Sutton and McCallum (2011:269)
Clustering [Unsupervised]	Parallelisation MapReduce k-nearest neighbour	A method where two methods are applied to the existing k-nearest neighbour (KNN) algorithm. Firstly, a clustering algorithm is used to group similar data to reduce the number of samples the KNN algorithm needs to process. Secondly, a map and reduced parallel model are applied to the data set to identify the independent categories of data and run the KNN algorithm multiple times simultaneously (in parallel). This improves the performance of the algorithm on larger samples	Du (2017)
	Semi-supervised clustering	Clustering algorithms are unsupervised ML algorithms that work to find a partition in the data set. Semi-supervised clustering assists the algorithm in finding a better-quality partition by providing the algorithm with any prior knowledge about the data. The clustering algorithm is then guided by the prior knowledge to find the partition in the data	Jain <i>et al.</i> (2015)
	K-means clustering	K-means clustering divides the data into k number of categories. To perform k-means clustering, the number of clusters, that is k, needs to be specified, and a random initial central data point (centroid) needs to be selected for each cluster. The data is then grouped based on the distance of each data point from the initial centre. The algorithm runs again until the cluster centres no longer need to move	Marsland (2009:196), Ayodele (2010b:27)

(continued)

ML solution	Algorithm	Description	Source
	Random forests	This model consists of a number of decision trees, each composed of a subsample of features, and is usually weaker than a full decision tree. The average, or the weighted average, of the trees is determined and used to perform the classification, effectively combining the power of the individual trees, which often produces a higher-quality result	Bucheli and Thompson (2014:4), Dataiku (2017:7)
	Self-organising maps	A form of neural network that uses unsupervised learning. The objective of the self-organising map is to produce its own representation or self-organisation of the given data as outputs are not provided	Kohonen (1990:1464), Hadzic <i>et al.</i> (2007:225)
Outlier detection [Unsupervised]	Association rules	Association rules determine the associative relationships between data, where the occurrence of one feature may indicate the possible occurrence of another feature. Instead of predicting a particular class, association rules can predict combinations of features and which features are commonly associated with each other, irrespective of class	Narasimha Murty and Susheela Devi (2011), Witten <i>et al.</i> (2016)

Source: Compiled by author from multiple sources as indicated above

Table A1.