Machine learning in predicting firm performance: a systematic review

China Accounting and Finance Review

Yaseen Hezam and Hoa Luong
Department of Accountancy and Finance, University of Otago, Dunedin, New Zealand, and
Lilian Anthonysamy
Faculty of Management, Multimedia University – Cyberjaya Campus, Cyberjaya, Malaysia

Received 24 March 2024
Revised 21 September 2024
5 December 2024
Accepted 3 January 2025

Abstract

Purpose – This study investigates the application of machine learning (ML) techniques in predicting firm performance, responding to the challenges posed by the large volumes of data required for accurate predictions. It aims to assess the effectiveness of various ML methods and algorithms used in recent research, focusing on the prediction of firm performance across multiple dimensions.

Design/methodology/approach – A systematic literature review was conducted, examining 70 studies published over the last decade (2013–2023) that utilize ML techniques for firm performance prediction. This methodology allowed for an in-depth analysis of the attributes, methods, and algorithms commonly applied in the field, offering insights into the evolution and effectiveness of these approaches over time.

Findings – The research highlights the importance of considering a broad range of attributes beyond traditional financial metrics, such as financial health, market positioning, operational efficiency, innovation capability, leadership quality, and employee engagement, in predicting firm performance. It reveals a predominance of classification methods in ML, with neural networks, logistic regression, and decision trees being the most frequently employed algorithms. These findings underscore the potential of ML techniques to provide a more nuanced and accurate prediction of firm performance by integrating diverse data sources and attributes.

Practical implications – The study's insights have significant implications for investors, financial analysts, corporate management, policymakers, and regulators. By adopting a more comprehensive ML-based approach to performance prediction, these stakeholders can make more informed decisions regarding resource allocation, capital budgeting, investment strategies, and policy formulation. Improved predictability also aids in the development of more effective regulations and policies, benefiting the broader economic landscape.

Originality/value – This research contributes to the existing literature by systematically reviewing and synthesizing the application of ML techniques in firm performance prediction over a substantial period. It offers a consolidated view of the methods and attributes that are most effective in this context, highlighting the shift towards more complex and holistic approaches to understanding firm dynamics. This comprehensive overview provides valuable insights for future research and practice in the field of business analytics and performance prediction.

Keywords Machine learning, Firm performance, Systematic review, Predictive analytics, Business intelligence

Paper type Literature review

1. Introduction

Predicting and evaluating firm performance has been a key area of interest in both business and academic sectors. Firm performance serves as a critical measure of organization's success, offering insights into strategic decision-making, competitive advantage, and overall growth

Declaration of generative AI and AI-assisted technologies in the writing process: During the preparation of this work the author(s) used ChatGPT in order to improve the writing of the paper. After using this ChatGPT, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

China Accounting and Finance Review
Vol. 27 No. 3, 2025
pp. 309-339
Emerald Publishing Limited
e-ISSN: 2307-3055
p-ISSN: 1029-807X
DOI 10.1108/CAFR-03-2024-0036

potential (Richard, Devinney, Yip, & Johnson, 2009; Venkatraman & Ramanujam, 1986). For organizations, especially in the corporate sector, understanding performance dynamics helps attract investors, engage clients, and identify opportunities for expansion (Giap *et al.*, 2021; Vukovic, Spitsina, Gribanova, Spitsin, & Lyzin, 2023). Moreover, firm performance is closely tied to strategic decision-making speed, which impacts growth and competitiveness in rapidly changing business environments (Robert Baum & Wally, 2003).

Operating in today's complex and competitive market, firms face multiple challenges, including providing high-quality services to their clients, establishing robust performance evaluation systems, and forecasting market demands (Boron, Villarba, Murcia, & Delima, 2023). Researchers are particularly interested in predicting firm performance by leveraging diverse variables such as financial performance, market position, operational metrics, the innovation capability, leadership, and employee engagement (Hajek, Olej, & Myskova, 2014; Chaithanapat, Punnakitikashem, Khin Khin Oo, & Rakthin, 2022; Khalil, Aziz, Long, & Zhang, 2023). The intensity of interest that firms have in their financial growth varies (Cozgarea, Cozgarea, Boldeanu, Pugna, & Gheorghe, 2023), where some firms can independently strategize using data analytics and technology, while other firms might need assistance of experts who design strategies to specific needs of a company (Hezam, Anthonysamy, & Suppiah, 2023; Sharma, Gupta, Sehrawat, Jain, & Dhir, 2023). However, the technological advancements, particularly in big data, have transformed the evaluation process, making it more intricate due to the vast amount of data and the dynamic factors influencing performance (Heredia *et al.*, 2022; Saito, Ohsato, & Yamanaka, 2021).

Big data, characterized by its volume, variety, and velocity, has emerged as a significant driver in predicting firm performance (Chen, Chiang, & Storey, 2012). Traditional indicators such as operational efficiency and leadership quality, while foundational, may fail to capture the complexities of modern business environments (Porter, 1985). By integrating real-time data on consumer behaviour, market dynamics, and operational processes, big data analytics offers opportunities for uncovering nuanced patterns that enhance predictive accuracy (McAfee & Brynjolfsson, 2012). This shift underscores the transformative potential of machine learning (ML) in augmenting traditional methods, allowing for more granular and robust evaluations (Kambatla, Kollias, Kumar, & Grama, 2014).

The motivation for this research stems from the need to systematically assess the intersection of big data and ML in firm performance. Unlike conventional studies that focus on building predictive models, this paper adopts a broader perspective by systematically reviewing existing research on the application of ML techniques in the business domain. This approach not only highlights the opportunities and challenges inherent in big data but also provides actionable insights into integrating diverse data sources for enhanced prediction accuracy.

Despite the proliferation of studies on firm performance and ML, there remains a lack of systematic research that consolidates findings on how different variables and ML algorithms influence prediction accuracy (Luong, Kumar, & Lang, 2020). This gap is particularly significant given the diversity of data and computational methods employed across studies. Addressing this gap, the present study aims to identify and examine key characteristics and ML techniques used in predicting firm performance, thereby contributing to both theoretical understanding and practical applications in the corporate sector.

To achieve this, a systematic literature review was conducted, encompassing studies from the past decade (2013–2023). The findings emphasize the prevalence of financial performance indicators and classification methods, with a strong focus on supervised learning techniques. This review not only consolidates current knowledge but also identifies opportunities for future research, particularly in leveraging diverse data sources to enhance ML model precision.

In practical terms, this study provides a framework for integrating traditional performance indicators with advanced ML techniques, offering valuable insights for decision-makers seeking to predictive capabilities and strategic planning. By bridging the gap between traditional methods and big data analytics, this research aims to guide firms in navigating the complexities of modern business environments effectively.

The structure of the paper is as follows: Section 2 reviews the related literature on machine learning and firm performance prediction. Section 3 outlines the research methodology, Section 4 presents the findings and discussion, and Section 5 discusses the future research directions and implications from theoretical and practical perspectives.

2. Research background

2.1 Firm performance

Firm performance is a pivotal concept in business studies, often used to evaluate the success of companies within the corporate sector (Mikalef, Boura, Lekakos, & Krogstie, 2019). Furthermore, it serves as a crucial indicator for organizations, reflecting their achievements and contributing to opportunities for expansion and growth and attracting investors and clients (Omar & Zineb, 2019). Firm performance is influenced by a multitude of factors including financial health (Aguilera, De Massis, Fini, & Vismara, 2024), market positioning (Arora & Brintrup, 2021), operational metrics (Handoyo, Suharman, Ghani, & Soedarsono, 2023), innovation capability (Schäper *et al.*, 2023), leadership (Chen, Sharma, Zhan, & Liu, 2019; Khan, Rehmat, Butt, Farooqi, & Asim, 2020), and employee engagement (Li, Naz, Khan, Kusi, & Murad, 2019). In today's competitive business environment, firms face challenges in providing quality services, establishing performance evaluation systems, and anticipating future market demands (Rožman, Tominc, & Štrukelj, 2023). The assessment of firm performance has evolved, going beyond traditional financial metrics to include a broader range of indicators. This comprehensive approach helps organizations achieve optimal performance, align with their mission, and meet societal and employee interests.

Financial health is a cornerstone of firm performance, which encompasses various aspects such as revenue growth, profitability, cash flow management, and asset utilization (Gleißner, Günther, & Walkshäusl, 2022). Financial stability is often the primary measure of a company's health, reflecting its ability to generate profit, manage debts, and sustain operations (Laghari, Ahmed, & López García, 2023). Key financial ratios like return on assets, return on equity, and debt-to-equity ratios provide insights into a firm's financial status, influencing investor confidence and strategic decision-making. According to Onufrey and Bergek (2021), market positioning relates to a firm's competitive stance in its industry that involves strategies to create a distinctive place in the minds of consumers relative to competitors, which is achieved through a unique value proposition, brand differentiation, and market segmentation and its crucial for attracting and retaining customers, achieving market share, and defining the company's long-term growth trajectory.

Moreover, Liu, Wu, Zhong, and Liu (2020) mentioned that operational metrics are vital indicators of a firm's efficiency and effectiveness in managing its day-to-day activities, which include measures like inventory turnover, supply chain efficiency, production costs, and quality control metrics as high operational efficiency often leads to reduced costs, improved product quality, and customer satisfaction, which in turn positively impacts firm performance. Accordingly, innovation capability is increasingly becoming a determinant of firm performance and thus refers to a company's ability to develop new products, services, or processes that provide competitive advantages (Rajapathirana & Hui, 2018). According to Jin and Choi (2019) Firms that foster a culture of innovation are more likely to adapt to market changes, meet evolving customer needs, and sustain long-term growth and innovation metrics might include the rate of new product development, investment in R&D, and the number of patents held.

Additionally, leadership plays a pivotal role in shaping firm performance as effective leadership involves strategic vision, decision-making capabilities, and the ability to inspire and motivate employees (Khan *et al.*, 2020; Sonmez Cakir & Adiguzel, 2020), and their capacity to steer the company through challenges and capitalize on opportunities significantly impacts the firm's trajectory and success. In addition, employee engagement is as well a critical to firm performance as engaged employees are more productive, have lower turnover rates, and contribute positively to the company's culture and customer satisfaction (Ghani, Hyder, Yoo, & Han, 2023). Kossyva, Theriou, Aggelidis, and Sarigiannidis (2023) high levels of engagement are correlated with better financial performance, innovation, and customer loyalty. Therefore, in the context of ML for predicting firm performance, these factors become critical variables. ML models can analyse these diverse data points to predict future performance more accurately. For example, financial data can be used to forecast profitability trends, while employee engagement metrics might predict operational efficiency. ML algorithms can correlate these variables with performance outcomes, providing nuanced insights into what drives success in different industries and business models. A deep understanding of these factors is essential for accurately predicting firm performance. They are not only crucial for business leaders in strategic decision-making but also for ML models to analyse and predict future trends. The integration of these diverse aspects into predictive models offers a comprehensive view of a firm's potential trajectory, allowing for more informed and strategic decision-making. Research could focus on developing sophisticated ML models that can effectively integrate and analyse these diverse factors, providing a holistic view of firm performance prediction.

2.2 Traditional methods for predicting performance

Traditionally, the prediction of firm performance relies on a variety of statistical and econometric models. These methods typically include linear regression, time-series analysis, and panel data models (Greene, 2012). Linear regression has been commonly used to estimate the relationships between dependent and independent variables, often leveraging financial ratios, market indicators, and historical performance metrics (Fama & French, 1992). Time-series analysis allows researchers to forecast future firm performance based on past trends, accommodating seasonality and cyclical fluctuations (Box, Jenkins, Reinsel, & Ljung, 2015). Panel data models combine cross-sectional and time-series data to analyse firm performance across different entities and over time, providing a richer context for prediction (Hsiao, 2014).

While these traditional methods have been fundamental in firm performance prediction, they present several limitations. First, they generally assume a linear relationship between variables, which may not capture the complexities and non-linear interactions inherent in modern business environments (Wooldridge, 2010). Second, these methods often depend on a limited set of indicators, potentially overlooking critical factors such as real-time market fluctuations, consumer behaviour, and innovation capabilities (Brockwell & Davis, 2016). Additionally, traditional approaches may struggle with the increasing volume, variety, and velocity of data available in the era of big data (McAfee & Brynjolfsson, 2012).

The advent of machine learning addresses these limitations by providing more sophisticated tools that can handle complex, non-linear relationships and integrate a broader range of variables (Jordan & Mitchell, 2015). Machine learning models can process vast amounts of diverse data, uncovering patterns and insights that traditional methods might miss. This capability enhances the predictive accuracy and robustness of models, making them more adaptable to the dynamic nature of firm performance (Sun, Sun, & Strang, 2018).

2.3 ML for predicting performance

Machine Learning (ML) has become increasingly significant in predicting firm performance (Saha, Young, & Thacker, 2023). As a part of artificial intelligence, ML's objective is to acquire knowledge from complex datasets to anticipate future outcomes (Makridakis, Spiliotis, & Assimakopoulos, 2018). Thus, unlike traditional statistical methods, ML focuses on accurate predictions, playing a crucial role in understanding and forecasting firm performance (Kureljusic & Karger, 2023). Various ML techniques, including classification, regression, and clustering, are employed to predict firm performance (Sarker, 2021). These methods have been instrumental in analysing firm-related data to identify trends and predictions. For instance, classification methods are used to categorize data, regression to uncover relationships, and clustering for grouping related data points. Prominent algorithms in this domain include neural networks, logistic regression, decision trees, support vector machines, and random forests (Mikalef *et al.*, 2019). ML algorithms have been applied in diverse contexts such as credit risk assessment, bankruptcy risk evaluation, and portfolio management (Leo, Sharma, & Maddulety, 2019). They enable the replication of expert decision-making processes by observing input variables and linking them to past decisions, thereby aiding in planning and evaluation (Bitetto, Cerchiello, Filomeni, Tanda, & Tarantino, 2023). The integration of ML in predicting firm performance represents a paradigm shift from traditional methods, which involves a holistic approach that considers both quantitative and qualitative dimensions (Rai, Tiwari, Ivanov, & Dolgui, 2021). This comprehensive understanding is vital for more accurate and predictive insights into firm performance and strategies for sustainable growth.

The exploration of predictive features in firm performance assessment underscores the evolving landscape of predictive models (Pap, Mako, Illessy, Kis, & Mosavi, 2022). It signifies a transition from a limited focus on financial metrics to a more holistic approach, encompassing various dimensions like market positioning, operational efficiency, innovation capability, leadership quality, and employee engagement (Leso, Cortimiglia, & Ghezzi, 2023). The landscape of ML algorithms for predicting firm performance is diverse, with each methodology offering unique strengths (Bitetto *et al.*, 2023). Neural networks, for example, excel in recognizing intricate patterns, while logistic regression is appreciated for its interpretability (Goodell, Kumar, Lim, & Pattnaik, 2021). Decision trees offer clear delineations of the impact of various financial ratios, and ensemble methods like random forests enhance prediction accuracy by capturing complex relationships (Ahmed, Alshater, Ammari, & Hammami, 2022). Enhancing the predictability of firm performance has significant theoretical and practical implications (Bargagli-Stoffi, Niederreiter, & Riccaboni, 2021). It can advance accounting theory, inform policymakers, and empower business leaders and stakeholders in decision-making. Predictive models help in strategic management and resource allocation, contributing to avoiding financial distress and improving investment strategies (Boukherouaa, AlAjmi, Deodoro, Farias, & Ravikumar, 2021).

The integration of ML in predicting firm performance is a dynamic and evolving field (Henrique, Sobreiro, & Kimura, 2019). It not only emphasizes the importance of financial metrics but also brings to light the significance of various other attributes that contribute to a firm's success (Reis, Ruivo, Oliveira, & Faroleiro, 2020). As the business environment continues to be dynamic, the ability to predict firm performance accurately using ML becomes increasingly vital (Chen, Esperança, & Wang, 2022). This research area encourages exploration of unconventional attributes and methodologies to enhance the precision of predictions, aiding decision-makers in shaping the future of their organizations. Future research could expand to incorporate external variables like economic conditions and regulatory changes for a more holistic understanding of firm performance prediction.

2.4 Summary and research gap

The literature emphasizes the shift in firm performance measurement from traditional financial indicators to holistic approaches, which include operational, innovative, and human capital perspectives. The traditional methods of firm performance prediction, though at the foundation of this approach, are constrained by linearity assumptions and limited integration of complex, real-time data. In contrast, the machine learning model, which allows the analysis of diverse, non-linear variables with dynamic datasets for more robust predictions. These developments underpin how ML has the capacity to overcome the deficiencies of the traditional models and provide a holistic platform for firm performance prediction.

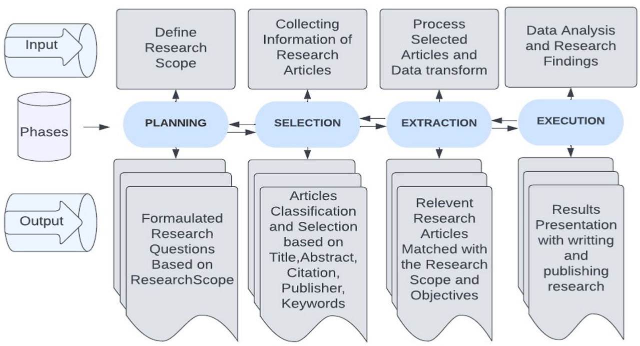
Despite these advancements, a critical gap persists in the application of ML for firm performance prediction. Notwithstanding these developments, an important gulf remains in the application of ML to predict firm performance. Previous studies have focused largely either on financial metrics or on the use of ML algorithms for narrowly bounded applications. Limited work has focused on integrated ML model development for holistic multidimensional variables involving leadership quality, employee engagement, and environmental externalities such as regulatory changes. This gap underlines the need for sophisticated ML frameworks that can synthesize heterogeneous data points into nuanced and predictive insights on firm performance. Capturing this gap will both enhance theoretical understanding and provide practical tools for decision-makers who navigate dynamic business environments.

This study aims to bridge this gap by proposing a comprehensive ML-based framework that incorporates diverse predictors, facilitating a more accurate and strategic decision-making in firm performance management.

3. Methodology

According to Brereton, Kitchenham, Budgen, Turner, and Khalil (2007), conducting a comprehensive systematic literature review entails employing an impartial study approach to ensure a consistency when it comes to reviewing all the relevant papers. The study follows three recommended stages, which are the review plan, the way to conduct the review, as well as the review reporting process as outlined in Kitchenham *et al.* (2010), as well incorporates with Okoli (2015) for doing a SLR. The methods produced a significant and standardized process for conducting a SLR, which despite being primarily focused on predicting firm performance using ML techniques, the study could be valuable and applicable to researchers in various scientific disciplines, including those related to business and finance. As shown in Figure 1, Okoli (2015) guidelines for conducting a systematic literature review.

A systematic structured review was employed to analyse related research studies on predicting firm performance. As indicated by Kitchenham *et al.* (2010), the initial phase of any review involves planning. Therefore, it is a necessary to establish and formulate a procedure plan for the review process. Our main goal in conducting this systematic literature review (SLR) is to assess its significance, and for this purpose, we have set the following objectives which are to identify the characteristics that serve as an indicator in predicting a firm performance and once these characteristics are established, the next objective will be to explore the techniques of ML used by researchers in predicting performance. Furthermore, it's crucial to determine which algorithms are most employed in the field of predicting firm



Source(s): Figure by authors

Figure 1. Systematic framework of literature review development

performance. The development of SLR questions is a critical step in the review methodology (Butler *et al.*, 2017). According to Kitchenham *et al.* (2010), it is essential to consider the population, intervention, outcomes, and context (PIOC) key points, which helps in formulating the research questions. Research questions criteria are detailed in Table 1. Referring to the table, the primary SLR question revolves around identifying the studies coverage in relation to firm performance and the appropriate ML methods used for predicting firm performance. In response to this query, the following specific questions were developed: What are the characteristics used when predicting firm performance? And which machine learning techniques are employed when predicting firm performance? Lastly, what are the effective algorithms and techniques in predicting a firm performance?

To achieve the objectives and outcomes related to predicting firm performance, it is crucial to use credible sources of information and search queries that are well-planned. A comprehensive investigation was conducted to address the research questions. Different search queries, such as wild card operators, key terms and Boolean operators were developed and implemented for ensuring that the relevant articles on firm performance prediction are extracted, as shown in Table 2.

Selecting papers aimed to identify studies related to firm performance and the selection process involved the identification, screening, verification of suitability, and satisfaction of inclusion criteria for research papers. Studies that investigated methods of predicting firm performance in different industries were also included.

The literature for this systematic review was retrieved from databases such as Scopus and Web of Science. Keyword searches were conducted on both article titles and abstracts using terms related to machine learning and firm performance. Inclusion criteria focused on peer-reviewed journal articles published between 2013 and 2023, while exclusion criteria filtered out non-peer-reviewed articles and those not directly addressing the research questions. The initial screening of studies was based on their publication titles and abstracts. Subsequently, the studies were further evaluated through a thorough examination of the full-text versions. In cases where there were uncertainties, studies were also considered for a comprehensive full-text review.

Table 1. PIOC criteria for SLR questions

Criteria	Detail
Population	Firms, companies, and organizations in the business domain and financial sector
Intervention	Methods, algorithms, and techniques employed in predicting firm performance
Outcome	Best performance indicators, key characteristics or factors influencing firm performance, as well as effective prediction strategies or approaches
Context	Business and corporate sector, and overall firm performance in the market

Source(s): Table by authors

Table 2. Search strings for the SLR

Research search	Keywords
Machine learning	"Machine learning" OR "machine learning algorithms" "Machine learning techniques" OR "neural network" OR "logistic regression" OR "naïve bayes" OR "support vector" OR "k-nearest" OR "linear regression" OR "decision trees" OR "random forest" OR "classification" OR "regression" OR "clustering" OR "data mining techniques" OR "data mining algorithms"
Firm	Firm* OR enterpris* OR compan* OR organiz* OR business*OR corporat*
Performance	performance OR failure OR success* OR achieve* OR bankruptcy* OR distress*
Prediction	Predict* OR forecast* OR assess* OR measur*

Source(s): Table by authors

In our review, we focused on the inclusion and exclusion criteria to ensure the selection of research publications that meet our objectives. We included research publications that primarily revolved around the predictions of firm performance and those that harnessed ML or data mining methodologies to predict firm performance. Additionally, we considered English written research, including those with English equivalent translations. Our emphasis was on research papers that underwent rigorous peer-review processes and were subsequently published in reputable journals. On the contrary, we employed exclusion criteria to maintain the quality and relevance of our selection. Research that did not incorporate ML or data mining techniques was excluded. We also omitted papers with duplicate entries, ensuring the uniqueness of our chosen literature. Furthermore, any research with unrelated headings, abstracts, or content was not included in our analysis. We excluded papers that solely relied on suggested methods without conducting experiments. Finally, we did not capture methodological research guidelines or handbooks with a broad focus, especially those that did not incorporate data mining or ML methodologies. These criteria were essential in streamlining our literature selection to best serve our research goals and objectives. The primary objective of the criteria for inclusion and exclusion was to encompass a broad range of publications relevant to the research objectives on firm performance. Criteria were specified and set to ascertain if the paper needed to be included in the analysis or excluded.

A compiled list of studies was extracted for further review and endnote was used for the purpose of keeping a record of the references, such as removing repetitive studies, as well as knowing the publication years. The bibliographies included in the research paper meet the specified requirements of the inclusion. As for the duplication screening, 1,287 publications were removed. For the screening criteria, titles and abstracts were fully screened that resulted in the removal of 3,648 publications, 32 publications were excluded with fully text screening as well 4 were eliminated according to a comprehensive full-text and content analysis which all the steps and procedures descripted in Figure 2.

4. Descriptive analysis

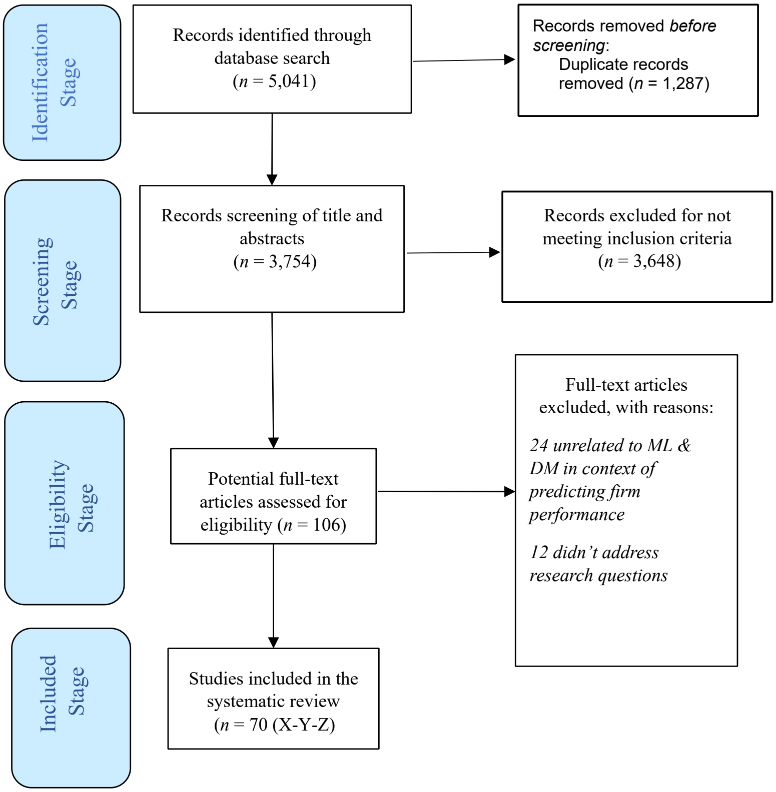
4.1 Descriptive analysis of papers

The overall publications captured 70 papers. Figure 3 depicts the year distribution of the included articles. According to the graph, following a significant decrease in published publications on company performance prediction from 2013 to 2018, the number of articles climbed considerably from 2019 to 2023.

This demonstrates that many firms have recently been interested in utilizing ML algorithms to forecast firm performance. According to Figure 3, the bulk of the included papers were published between 2020 and 2023.

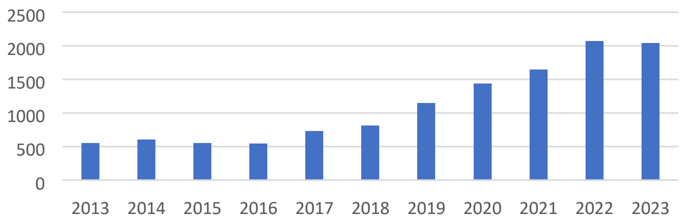
It is noteworthy that a substantial proportion of research papers in this field, as shown in Figure 4 originates from countries such as the US, China, India, and the UK. In this collaborative and interdisciplinary research paradigm, the confluence of expertise across diverse domains augments the efficacy of ML applications, facilitating nuanced insights and a comprehensive understanding of the intricate dynamics governing firm performance. In addition, the prominence of research papers on ML and firm performance from the US, China, India, and the UK can be attributed to a combination of strong research infrastructure, economic significance, industry collaboration, global competitiveness, government support, a talented workforce, and active participation in international research networks. These factors collectively contribute to the leadership role of these countries in advancing knowledge and applications in this field.

The paradigm of employing machine learning (ML) for predicting a firm performance underscores the imperative of a synergistic and interdisciplinary approach, wherein expertise is disseminated across multifarious domains. Notably, Figure 5 shows that the discipline of computer science assumes a foundational position, commanding a substantial 33.3% influence. Across the entire ML model developmental continuum, experts in computer science



Source(s): Figure by authors

Figure 2. Flow diagram of papers included in review based on Prisma (2020)

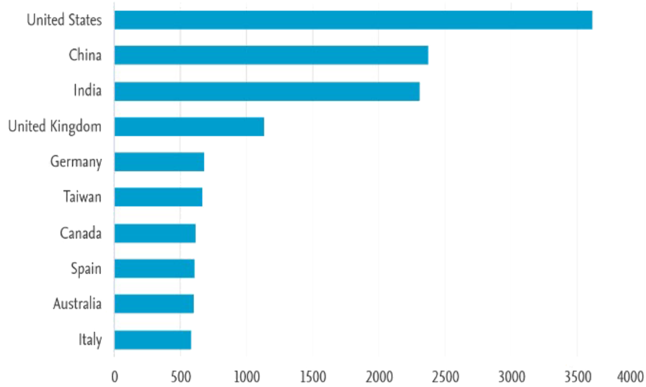


Source(s): Figure by authors

Figure 3. Yearly published research on ML performance prediction

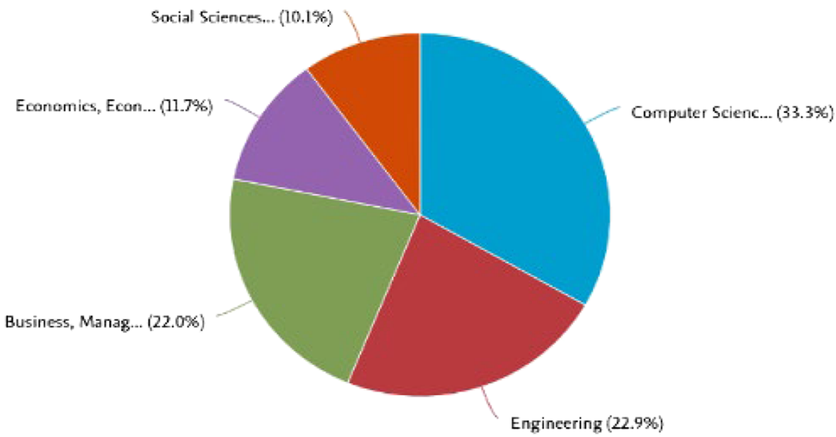
are pivotal agents in executing tasks spanning data preprocessing, algorithmic selection, model training, testing, and deployment. Their purview extends towards the conception and refinement of models adept at processing expansive datasets, incorporating financial metrics, market trends, and other salient indicators.

Concomitantly, engineering disciplines, constituting 22.9%, wield critical influence in the augmentation of firm performance prediction through ML applications. Specifically, within



Source(s): Figure by authors

Figure 4. Research distribution by country



Source(s): Figure by authors

Figure 5. Subject distribution

the industrial, mechanical, and systems engineering domains, professionals adeptly employ ML methodologies in optimizing manufacturing processes, managing supply chains, and facilitating product development. The consequent optimization of operational paradigms and precision in predicting maintenance schedules indubitably resonate in the overall enhancement of firm performance.

Simultaneously, the cohort of business management researchers, wielding a substantive 22.0% influence, assumes a pivotal role in harnessing ML outputs to inform strategic decision-making. Their ability lies in the nuanced interpretation of model predictions within the broader business environment, encompassing considerations of market conditions, competitive dynamics, and internal corporate metrics. Furthermore, their insights extend towards explicating the ramifications of diverse business strategies on predictive models, thereby amplifying the pragmatic utility and applicability of predictions.

In addition, economists, representing 11.7% of the collaborative endeavour, contribute significantly to firm performance prediction through the judicious integration of ML techniques for the analysis and prediction of market trends, demand oscillations, and economic indicators. The incorporation of macroeconomic and microeconomic data into ML models is instrumental in engendering a holistic understanding of the contextual determinants impacting firm performance predictions.

Conclusively, the social sciences discipline, with a 10.1% share, imparts invaluable insights into consumer behaviour, societal trends, and employee satisfaction. This qualitative dimension enriches ML models tailored for firm performance prediction, imbuing them with the capacity to encapsulate non-quantitative factors of paramount significance such as brand perception and customer loyalty. In this collaborative and interdisciplinary research paradigm, the confluence of expertise across these diverse domains' manifests in an augmented efficacy of ML applications, facilitating nuanced insights and a comprehensive comprehension of the intricate dynamics governing firm performance.

5. Discussion and findings

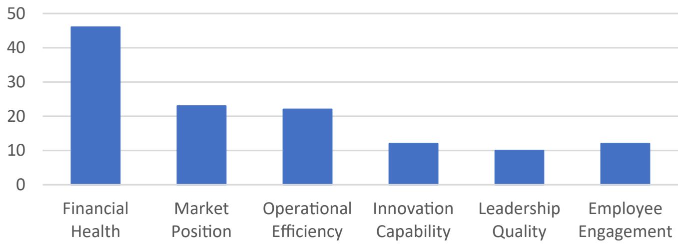
The section investigates and covers the features as well as the ML methods and techniques used in this review study of firm performance prediction. Factors that encompass the qualities and traits will determine the accuracy of data sources and techniques when it comes to predicting firm performance, which will be used for the purpose of analysing firm-relevant data to identify trends and predictions. The results of the analysis are expected to fulfil the objectives of the review.

5.1 Performance prediction

Numerous studies have unveiled a diverse array of factors influencing the accuracy of firm performance prediction. These factors may bear various labels; several share similar attributes and can be amalgamated into unified categories. The importance of firm performance lies in its multidimensional nature, with financial performance being a prominent aspect (Gentry & Shen, 2010; Gupta, Rawte, & Zaki, 2023). For this review, specific variables were identified and compiled. This research literature revealed six overarching feature categories commonly discussed in the context of predicting firm performance, financial health, market positioning, operational efficiency, innovation capability, leadership quality, and employee engagement. Our analysis delved into 70 key papers relevant to firm performance prediction, identifying six overarching feature categories: financial performance, market positioning, operational efficiency, innovation capability, leadership quality, and employee engagement. Each of these categories reflects a different dimension of firm performance, illustrating the multifaceted nature of the subject. In our analysis, 46% of the studies utilized attributes from the financial performance category, making it the most frequently employed set of variables. Market positioning followed with a 23% prevalence, and operational efficiency was at 22%. The lower utilization rates for leadership quality (10%), innovation capability (12%), and employee engagement (12%) suggest that these areas may offer unexplored potential in firm performance prediction. This discrepancy highlights the dominant focus on financial metrics in the literature while also pointing to the value of a broader analytical approach.

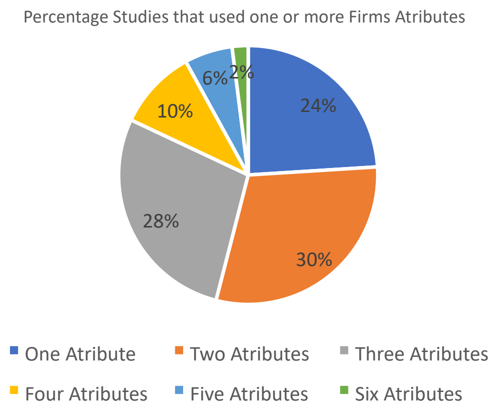
Figure 6 provides evidence that most researchers regard financial performance indicators, such as revenue growth, profit margins, and financial ratios, as pivotal determinants of firm performance.

Figure 7 indicates that employing multiple attributes enhances prediction accuracy. Notably, 30% of researchers in our review study incorporated two different sets of attributes in their analyses, with 28% utilizing three attributes and 10% employing four. Only 6 and 2% of the review studies investigated the leadership and employee attributes combinations, respectively, while 24% of the studies relied on a single attribute. The significance of



Source(s): Figure by authors

Figure 6. Attributes used for prediction of firm performance



Source(s): Figure by authors

Figure 7. Percentage attributes distribution

incorporating a multitude of characteristic categories cannot be overstated, as 76% of the authors employed multiple categories of characteristics, irrespective of their focus. Furthermore, as illustrated in Figure 4, performance indicators consistently demonstrate that financial metrics (46%) are the most frequently employed attributes.

Figures 6 and 7 provide visual evidence that financial performance indicators, such as revenue growth and profit margins, are pivotal determinants of firm performance. However, the integration of multiple attributes enhances prediction accuracy. Notably, 76% of researchers employed multiple categories of characteristics, indicating a trend toward more holistic predictive models.

Additionally, Table 3 highlights the attributes and their domains, which identified the profits and revenue growth as an optimal indicator for financial performance within the financial metric domain. Most studies concur that firm performance is significantly influenced by metrics such as financial health and market position. Market positioning features, despite a relatively high recurrence rate of 23%, still lag financial health features in terms of utilization. However, as highlighted in Table 3, researchers have noted that market positioning, including market share, customer base, and competitive analysis, holds substantial predictive power. While leadership quality features have received less attention in the literature, our evaluation identified 10 studies emphasizing the importance of factors like qualification and track record in predicting firm performance. In the realm of firm performance prediction, it is crucial for researchers to explore unconventional attributes beyond traditional metrics. Factors such as

Table 3. Attributes frequently used in predicting firm performance

Attributes	Attributes domain	Frequency
Financial performance	This attribute can include metrics such as revenue growth, profit margins, and financial ratios (e.g. liquidity, solvency)	46
Market position	Evaluate factors like market share, customer base, and competitive analysis	23
Operational efficiency	Assess factors like production efficiency, supply chain effectiveness, and cost management	22
Innovation capability	Consider research and development spending, patents, and innovation awards	12
Leadership quality	Include attributes related to the leadership team, such as experience, qualifications, and track record	10
Employee engagement	Measure factors like employee satisfaction, turnover rate, and skill development	12

Source(s): Table by authors

market positioning, operational efficiency, innovation capability, leadership quality, and employee engagement offer valuable avenues for deeper insights. These often-overlooked aspects, analogous to indicators in other domains, may hold the key to enhancing the accuracy of predictions.

The review, structured around these six feature categories as outlined in Table 3, underscores the importance of considering a broad spectrum of characteristics. This comprehensive analysis demonstrates the intricate interplay of multifaceted attributes shaping firms' trajectories. Traditional financial metrics undoubtedly wield substantial influence, but this study emphasizes the imperative of transcending these conventional indicators to capture a holistic perspective of firm performance.

The assessment of firm performance extends beyond traditional financial metrics, encompassing a broad spectrum of predictive features that provide a more comprehensive understanding of what drives success. This study delves into these variables, as illustrated in Table 3, revealing the diverse elements that influence and forecast a firm's trajectory. This analysis demonstrates that while financial indicators such as revenue, profit margins, and return on assets remain critical, they represent only one facet of a more complex picture. Traditional financial metrics undoubtedly wield substantial influence, and this study accentuates the imperative of transcending these conventional indicators to capture a holistic perspective of firm performance.

This recognition marks a pivotal evolution in the predictive modelling of firm performance. Traditional financial indicators have long dominated the predictive landscape due to their quantifiable nature and direct impact on performance outcomes. However, overlooking other the reliance of these metrics alone can lead to an incomplete and potentially misleading assessments of a firm's health and prospects. The inclusion of non-financial dimensions represents a paradigm shift, broadening the scope of predictive models to encompass a richer and more nuanced understanding of what drives firm success.

The depth of understanding required to effectively integrate these varied attributes into predictive frameworks paves the way for more inclusive models. These models, by accounting for both quantitative and qualitative dimensions, can offer a more accurate and robust representation of firm performance. This approach not only improves the predictive power of these models but also aligns with the increasing complexity of the business environment, where success is often determined by a confluence of financial strength, innovative capacity, and effective leadership.

Furthermore, the call for a balanced consideration of financial and non-financial facets resonates with broader research across diverse industries and regions. Smith *et al.* (2022) argued for the integration of qualitative dimensions, such as employee engagement and innovation capacity, alongside traditional financial metrics. Their empirical investigations underscore the pivotal role these non-financial attributes play in shaping a firm's long-term trajectory and the competitive advantage.

This evolving understanding of predictive features in firm performance assessment signifies a transition from a narrow focus on financial metrics to a more holistic and inclusive approach. Embracing a comprehensive understanding of attributes encompassing both quantitative and qualitative dimensions, it enables predictive models to provide more nuanced insights into firm performance. These insights are crucial for developing strategies that not only drive growth but also ensure sustainability in an increasingly complex and dynamic business landscape. This perspective is reinforced by Smith *et al.* (2022), who emphasize the inclusion of qualitative dimensions such as employee engagement and innovation capacity for a thorough evaluation of firm performance.

5.2 Machine learning techniques for predicting firm performance

According to the research question, which asks which ML techniques are most frequently used to forecast firm performance, a thorough review of the literature reveals that three primary ML techniques classification, clustering, and regression are frequently employed in predicting firm performance. This is reflected in Figure 6.

While regression is employed to uncover relationships between dependent and independent variables, enabling researchers to understand how changes in predictors affect outcomes, classification techniques aim to accurately categorize data into specific classes (Roumani, Nwankpa, & Tanniru, 2020). For example, classification methods can be used to categorize firms into performance tiers based on various metrics, facilitating targeted analyses and interventions. In contrast, clustering represents an unsupervised learning approach distinctly from both regression and classification. Clustering does not rely on predefined labels; instead, it groups related data points based on inherent similarities within the dataset (Park, Son, Hyun, & Hwang, 2021). This approach is particularly useful for discovering underlying patterns or segments within data that are not immediately apparent, offering valuable insights into the structure and dynamics of firm performance. Our review indicates that classification methods are the most frequently applied techniques for predicting firm performance. Specifically, classification methods are utilized in 43 instances for financial performance predictions and 24 instances for market position predictions. The high usage reflects the effectiveness of classification in managing and interpreting labelled data, which is crucial for accurate performance prediction. The prevalence of classification techniques is attributed to their simplicity and computational efficiency. As supervised learning methods, they utilize labelled datasets to train algorithms, allowing them to learn and categorize data or predict outcomes with a high degree of accuracy (Loh, 2011). This makes classification techniques a preferred choice for researchers seeking reliable and interpretable results (refer to Table 4 for a detailed comparison).

Conversely, clustering is the second most frequently used technique, with 32 instances applied to financial health predictions and 17 to market position predictions.

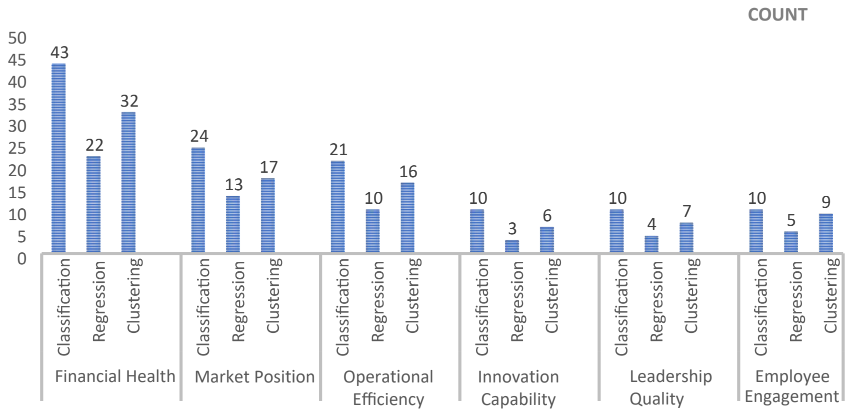
Although less common than classification, clustering plays a crucial role in uncovering latent data structures and identifying patterns that may not be visible through supervised methods. This capability makes clustering particularly valuable for exploratory data analysis and identifying new trends or segments within firm performance data.

Figures 6 and 8 illustrates the distribution of these ML techniques in firm performance prediction, highlighting the dominance of classification methods. This figure underscores the preference for classification due to its straightforward application and the ability to leverage labelled data for precise predictions. In light of this, Brown & Smith (2022) reaffirm the pivotal role that supervised learning plays in predicting firm performance, highlighting the capacity of this learning method to identify patterns and correlations within data, hence facilitating precise forecasts based on pre-existing patterns. The role of clustering, while secondary, remains

Table 4. Performance prediction methods

Features	Method	Count
Financial performance	Classification	43
	Regression	22
	Clustering	32
Market position	Classification	24
	Regression	13
	Clustering	17
Operational efficiency	Classification	21
	Regression	10
	Clustering	16
Innovation capability	Classification	10
	Regression	3
	Clustering	6
Leadership quality	Classification	10
	Regression	4
	Clustering	7
Employee engagement	Classification	10
	Regression	5
	Clustering	9

Source(s): Table by authors



Source(s): Figure by authors

Figure 8. ML technique distribution on firm performance

significant as it complements classification by offering deeper insights into data structure and relationships.

In summary, while classification methods are preferred for their efficacy and efficiency in predicting firm performance, clustering methods provide essential supplementary insights by revealing hidden patterns and relationships. Together, these techniques offer a comprehensive toolkit for understanding and enhancing firm performance through advanced data analysis.

In addition to classification techniques, this analysis extensively examines the role of clustering techniques, a prominent unsupervised learning approach, in deciphering intricate data patterns.

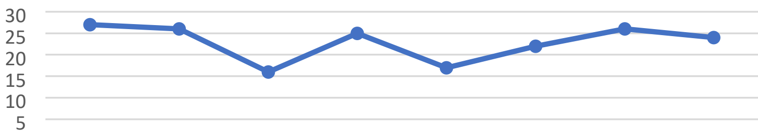
Regression methods are highlighted as the foundational elements of this analysis. Garcia & Martinez (2024) emphasize the significance of regression models in capturing both linear and nonlinear relationships between predictor variables and performance outcomes such as revenue and profitability. Regression techniques, including linear regression, polynomial regression, and generalized linear models, are instrumental in quantifying the strength and nature of relationships between variables. This capability provides valuable predictive insights by elucidating how changes in predictor variables affect performance outcomes. The use of regression techniques offers a complementary perspective to classification and clustering methods by providing a detailed examination of variable relationships and their impact on firm performance.

The integration of supervised methods, particularly classification with unsupervised clustering and regression techniques, highlights a methodologically rigorous approach to data analysis. Supervised learning, exemplified using classification techniques, leverages labelled datasets to train algorithms that predict outcomes with precision. In contrast, unsupervised learning through clustering techniques enables the exploration of data without predefined labels, thus revealing underlying patterns and structures. The inclusion of regression further enriches the analysis by addressing both linear and nonlinear relationships within the data. This structured approach acknowledges the multifaceted nature of data patterns, which may not be fully captured through labelled observations alone. By combining these methodologies, the analysis not only enhances the predictive accuracy but also provides a comprehensive understanding of the diverse dimensions of firm performance. This holistic approach ensures that insights are drawn from multiple analytical perspectives, thereby strengthening the overall robustness of the predictive model. This sequential exploration of classification, clustering, and regression methodologies presents a comprehensive and integrated approach to understanding and predicting firm performance. The comprehensive framework emphasizes the importance of utilizing a range of techniques to capture the full complexity of data and generate precise, actionable insights from multifaceted data sources.

5.3 Most used algorithms

In addressing the research question regarding the most commonly employed algorithms for predicting firm performance, this review presents a comprehensive examination of various machine learning (ML) techniques. This analysis reveals a diverse array of algorithms, indicating the broad range of approaches utilized in predictive modelling within the domain of firm performance. Figure 9 illustrates the distribution of these algorithms across reviewed studies, highlighting the prevalence and application of each method, while Table A3 (Appendix) provides a detailed breakdown of the references and citation counts for each algorithm.

Figure 9 shows that the random forest (RF), decision tree (DT), artificial neural network (ANN), Bayesian networks (BN), logistic regression (LR), K nearest neighbour (KNN), and support vector machine (SVM) are notable methods that scholars commonly employ to



	Neural Network	Logistic Regression	Naïve Bayes	Support Vector	k-Nearest	Linear Regression	Decision Trees	Random Forest
Count	27	26	16	25	17	22	26	24

Source(s): Figure by authors

Figure 9. Machine learning algorithms used for firm performance

forecast firm performance. As detailed in Table A3, among the algorithms frequently employed, neural networks (NN) algorithm is notably prominent, with 27 references in the reviewed literature. Neural networks are distinguished by their capacity to recognize complex patterns and capture nonlinear relationships within financial data. Their effectiveness in handling intricate datasets makes them a valuable tool for predicting firm performance, as they excel in identifying nuanced variations and interactions among financial metrics. The ability of neural networks to model intricate dependencies and adapt to dynamic data environments underpins their dominant role in predictive analytics (Athey & Imbens, 2019).

Further analysis of Figure 9 and Table A3 shows that logistic regression (LR) and decision tree (DT) emerge as the second most used algorithms, each with 26 references. Logistic regression is prized for its simplicity and interpretability, making it a robust choice for modelling binary outcomes such as firm performance categories. Its straightforward approach allows researchers to gain clear insights into the impact of various financial metrics on performance outcomes. For example, these trees segment data based on significant attributes, offering clear delineations of the impact of various financial ratios on firm performance. Similarly, decision trees offer an intuitive visualization of decision-making processes, segmenting data based on significant attributes and thereby providing a transparent understanding of how different financial ratios influence firm success.

Support vector machines (SVMs) and random forests (RF) also play significant roles in predictive modelling. SVMs are particularly adept at managing complex, nonlinear relationships within financial data, offering a powerful tool for classifying firms into distinct performance categories. The versatility of SVMs in handling high-dimensional data contributes to their relevance in performance prediction. Random forests, an ensemble learning method, further enhance predictive accuracy by aggregating predictions from multiple decision trees. This technique mitigate overfitting and improves the robustness of predictions by combining diverse insights from individual trees (Breiman, 2001).

Additionally, k-Nearest Neighbors (KNN) and Bayesian networks (BN) are notable for their specific contributions. KNN, a non-parametric technique, leverages similarity-based categorization to identify patterns among firms with similar financial characteristics, offering valuable insights into clustering and segmenting firms based on proximity in the feature space. Bayesian networks, on the other hand, incorporate probabilistic dependencies among variables, providing a framework for understanding and predicting the probabilistic relationships within financial data.

Linear regression, although a classic statistical method, retains significance in firm performance prediction (Vuković *et al.*, 2023). By modelling relationships between dependent and independent variables, linear regression unveils quantitative insights into how changes in financial metric impact firm outcomes. The advent of ensemble methods like random forests has further revolutionized predictive modelling. Random forests, comprising multiple decision trees, reduce overfitting and enhance prediction accuracy by aggregating diverse predictions from individual trees. This technique adeptly captures intricate relationships between various financial indicators and firm performance.

The choice of machine learning algorithm is influenced by several factors, including the complexity of the problem, the nature of the data, and the need for interpretability. Researchers select algorithms based on these considerations to address specific challenges and optimize prediction accuracy. This diverse methodological approach underscores the necessity of employing a multifaceted strategy to decipher the complex dynamics of firm performance.

As machine learning continues to evolve, the ongoing exploration of novel algorithms and their applications promises further advancements in predictive modelling. The integration of these methodologies highlights the need for a comprehensive and adaptable approach in firm performance analysis, ensuring accurate and insightful predictions from complex and varied data sources. The collective utilization of diverse algorithms illustrates the evolving landscape of predictive analytics and the critical role of multi-faceted approaches in enhancing the precision and relevance of firm performance predictions.

6. Conclusion

This systematic literature review provides an extensive examination of machine learning applications in predicting firm performance, offering both theoretical and practical insights into prevalent attributes, methods, and algorithms. This review underscores the dominance of financial metrics and classification methods, highlighting the critical role of supervised learning in producing precise predictions. Commonly employed algorithms like neural networks, logistic regression, and decision trees are noted for their adaptability and efficacy in various scenarios. Additionally, support vector machines and random forests are acknowledged for their unique contributions, particularly in handling complex and nonlinear data relationships. The selection of algorithm remains context-dependent, varying according to the specific characteristics of the dataset and the research objectives. As businesses navigate increasingly volatile and complex environments, the ability to predict firm performance becomes more crucial. This review lays a foundational framework for further scholarly inquiry and encourages the integration of unconventional attributes and methodologies to refine predictive accuracy. By fostering a broader exploration of diverse data dimensions, researchers can provide decision-makers with more nuanced tools to anticipate and influence the trajectory of their organizations However, this review primarily focuses on machine learning techniques, potentially overlooking significant external factors influencing firm performance predictions. Elements such as macroeconomic conditions, regulatory shifts, and geopolitical events, while critical, are not thoroughly examined within the scope of this study. Addressing these limitations in future research will be essential for developing a more holistic understanding of the determinants of firm performance, thereby enhancing the robustness and applicability of predictive models in real-world business contexts. Furthermore, future reviews should consider incorporating articles from the UT24 and FT50 databases to ensure a more comprehensive coverage of relevant studies, particularly those that may provide additional insights into the relationship between machine learning and firm performance.

7. Future research

7.1 Investigating non-traditional predictors

Future research should extend beyond traditional financial metrics to investigate a wider array of predictors. This could include exploring non-financial indicators and qualitative data, such as customer sentiment, social media interactions, and environmental, social, and governance (ESG) factors. By integrating these novel features, researchers can develop predictive models that capture a broader spectrum of dynamics influencing firm performance.

7.2 Adapting to dynamic business environment

Given the volatile business landscapes, there is a need for predictive models that can adjust to changing conditions. Future studies should focus on developing models that can dynamically respond to economic fluctuations, market shifts, and other external factors, thereby enhancing the resilience and accuracy of firm performance forecasts.

7.3 Cross- disciplinary perspectives

Incorporating insights from fields beyond finance and machine learning, such as economics, psychology, or sociology, can provide a more holistic understanding of factors affecting firm performance. By adopting a cross-disciplinary approach, researchers can identify hidden patterns and relationships, leading to a more robust and comprehensive predictive models.

7.4 Temporal dynamics in prediction models

Exploring the temporal aspects of firm performance prediction is crucial for understanding how predictive models perform over different time horizons. Future research should assess the effectiveness of these models in capturing both short-term and long-term trends, thus facilitating the development of tailored forecasting tools that meet specific temporal requirements.

7.5 Integration of external influences

To address current limitations, future research should focus on incorporating external factors such as economic indicators, regulatory changes, and geopolitical events into ill predictive models. This integration will enhance the accuracy and robustness of predictions by providing a more comprehensive view of the variables that impact firm performance.

7.6 Enhancing model explainability

As machine learning models become more complex, ensuring their explainability and interpretability becomes increasingly important. Future research should aim to develop techniques that make these models more transparent and understandable, thereby building trust among stakeholders and ensuring their practical application in real-world business scenarios. These proposed directions for future research aim to refine and expand the current understanding of firm performance prediction, ensuring that the models developed are both accurate and adaptable to the evolving business environment.

8. Theoretical implications

Enhancing the predictability of firm performance holds substantial theoretical significance. By providing deeper insights into the drivers of economic growth, market dynamics, and corporate behaviour, this advancement contributes significantly to the evolution of accounting and financial theories. Understanding and predicting financial risk factors, in particular is an important area where improved predictability can lead to the development of more sophisticated risk models. These models not only deepen the theoretical understanding of market fluctuations and financial risk management but also enable more robust testing and validation of existing financial theories. Furthermore, this improved predictability offers a theoretical framework for policymakers and regulators to devise more effective regulations and policies. By accurately identifying the factors that influence firm performance, researchers and policymakers can collaborate to develop targeted interventions that support struggling firms and to foster an environment conducive to business growth. This alignment between theoretical insights and regulatory practices underscores the broader societal impact of enhanced predictability in firm performance.

9. Practical implications

From a practical perspective, improved predictability of firm performance equips business leaders, investors, and stakeholders with the tools to make more informed decisions. This empowerment leads to optimized resource allocation, strategic planning, and risk mitigation. Firms can utilize predictive models to refine their strategic management approaches, adapting their strategies proactively in response to early warning signs and thereby potentially averting financial distress or bankruptcy. Furthermore, the accurate prediction of firm performance has profound implications for corporate finance and investment strategies. Investors, financial analysts, and corporate decision-makers can leverage these predictive insights to make more judicious decisions regarding resource allocation, capital budgeting, and investment choices. This practical utility not only enhances the financial stability and growth potential of individual firms but also contributes to the overall efficiency and resilience of financial markets. By fostering a culture of data-driven decision-making, enhanced predictability serves as a catalyst for innovation and sustainable growth in the corporate sector.

Reference

Aguilera, R. V., De Massis, A., Fini, R., & Vismara, S. (2024). Organizational goals, outcomes, and the assessment of performance: Reconceptualizing success in management studies. *Journal of Management Studies*, *61*(1), 1–36. doi: 10.1111/joms.12994.

Ahmed, S., Alshater, M. M., Ammari, A. E., & Hammami, H. (2022). Artificial intelligence and machine learning in finance: A bibliometric review. *Research in International Business and Finance*, *61*, 101646. doi: 10.1016/j.ribaf.2022.101646.

Alanis, E., Chava, S., & Shah, A. (2023). Benchmarking machine learning models to predict corporate bankruptcy. *The Journal of Credit Risk*, *19*(2), 77–110. doi: 10.21314/JCR.2023.002.

Arora, S., & Brintrup, A. (2021). How does the position of firms in the supply chain affect their performance? An empirical study. *Applied Network Science*, *6*(1), 19. doi: 10.1007/s41109-021-00364-9.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, *11*(1), 685–725. doi: 10.1146/annurev-economics080217-053433.

Awoin, E., Appiahene, P., Gyasi, F., & Sabtiwu, A. (2020). Predicting the performance of rural banks in Ghana using machine learning approach. *Advances in Fuzzy Systems*, *2020*, 8028019. doi: 10.1155/2020/8028019.

Bahrami, M., Boz, H. A., Suhara, Y., Balcisoy, S., Bozkaya, B., & Pentland, A. (2023). Predicting merchant future performance using privacy-safe network-based features. *Scientific Reports*, *13* (1), 10073. doi: 10.1038/s41598-023-36624-0.

Bargagli-Stoffi, F. J., Niederreiter, J., & Riccaboni, M. (2021). Supervised learning for the prediction of firm dynamics. In S. Consoli, D. Reforgiato Recupero, & M. Saisana (Eds.), *Data science for economics and finance: Methodologies and applications* (pp. 19–41). Springer International Publishing. doi: 10.1007/978-3-030-66891-4_2.

Ben Jabeur, S., Stef, N., & Carmona, P. (2023). Bankruptcy prediction using the XGBoost algorithm and variable importance feature engineering. *Computational Economics*, *61*(2), 715741. doi: 10.1007/s10614-021-10227-1.

Bitetto, A., Cerchiello, P., Filomeni, S., Tanda, A., & Tarantino, B. (2023). Machine learning and credit risk: Empirical evidence from small- and mid-sized businesses. *Socio-Economic Planning Sciences*, *90*, 101746. doi: 10.1016/j.seps.2023.101746.

Boron, V. R. O., Villarba, L., Murcia, J. V., & Delima, A. J. (2023). Applying data mining techniques to predict the market performance of publicly traded companies. *International Journal of Advanced and Applied Sciences*, *10*(7), 17–22. doi: 10.21833/ijaas.2023.07.003.

Boubaker, S., Le, T. D., Ngo, T., & Manita, R. (2023). Predicting the performance of MSMEs: A hybrid DEA-machine learning approach. *Annals of Operations Research*, 1–23.

Boukherouaa, E. B., AlAjmi, K., Deodoro, J., Farias, A., & Ravikumar, R. (2021). Powering the digital economy: Opportunities and risks of artificial intelligence in finance. *Departmental Papers*, *2021*(024), A001. doi: 10.5089/9781589063952.087.A001.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Hoboken, NJ: John Wiley & Sons.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi: 10.1023/A:1010933404324.

Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, *80*(4), 571–583. doi: 10.1016/j.jss.2006.07.009.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting* (2nd ed.). New York: Springer-Verlag.

Brown, K., & Smith, J. (2022). The role of quantitative finance in financial market analysis and decision-making. *Financial Analysts Journal*, *78*, 44–60.

Butler, M., Epstein, R. A., Totten, A., Whitlock, E. P., Ansari, M. T., Damschroder, L. J., . . . Guise, J.-M. (2017). AHRQ series on complex intervention systematic reviews—paper 3: Adapting frameworks to develop protocols. *Journal of Clinical Epidemiology*, *90*, 19–27. doi: 10.1016/j.jclinepi.2017.06.013.

Chaithanapat, P., Punnakitikashem, P., Khin Khin Oo, N. C., & Rakthin, S. (2022). Relationships among knowledge-oriented leadership, customer knowledge management, innovation quality and firm performance in SMEs. *Journal of Innovation and Knowledge*, *7*(1), 100162. doi: 10.1016/j.jik.2022.100162.

Che, S., Zhu, W., & Li, X. (2020). Anticipating corporate financial performance from CEO letters utilizing sentiment analysis. *Mathematical Problems in Engineering*, *2020*, 5609272. doi: 10.1155/2020/5609272.

Chen, W.-S., & Du, Y.-K. (2009). Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*, *36*(2, Part 2), 4075–4086. doi: 10.1016/j.eswa.2008.03.020.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165–1188. doi: 10.2307/41703503.

Chen, J.-X., Sharma, P., Zhan, W., & Liu, L. (2019). Demystifying the impact of CEO transformational leadership on firm performance: Interactive roles of exploratory innovation and environmental uncertainty. *Journal of Business Research*, *96*, 85–96. doi: 10.1016/j.jbusres.2018.10.061.

Chen, D., Esperança, J. P., & Wang, S. (2022). The impact of artificial intelligence on firm performance: An application of the resource-based view to e-commerce firms. *Frontiers in Psychology*, *13*, 884830. doi: 10.3389/fpsyg.2022.884830.

Colak, G., Fu, M., & Hasan, I. (2020). Why are some Chinese firms failing in the US capital markets? A machine learning approach. *Pacific-Basin Finance Journal*, *61*, 101331. doi: 10.1016/j.pacfin.2020.101331.

Cozgarea, A. N., Cozgarea, G., Boldeanu, D. M., Pugna, I., & Gheorghe, M. (2023). Predicting economic and financial performance through machine learning. *Economic Computation and Economic Cybernetics Studies and Research*, *57*(2), 89–102. doi: 10.24818/18423264/57.2.23.06.

Creamer, G., & Freund, Y. (2010). Using boosting for financial analysis and performance prediction: Application to S&P 500 companies, Latin American ADRs and banks. *Computational Economics*, *36*(2), 133–151. doi: 10.1007/s10614-010-9205-3.

Delen, D., Kuzey, C., & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications*, *40*(10), 3970–3983. doi: 10.1016/j.eswa.2013.01.012.

Dong, B., Wang, X., & Cao, Q. (2022). Performance prediction of listed companies in smart healthcare industry: Based on machine learning algorithms. *Journal of Healthcare Engineering*, *2022*, 8091383. doi: 10.1155/2022/8091383.

Elamir, E. A. (2021). Boosting algorithms to analyse firm's performance based on return on equity: An explanatory study. *International Journal of Computing and Digital Systems*, *10*(1), 1031–1041. doi: 10.12785/IJCDS/100193.

Eom, T., Woo, C., & Chun, D. (2022). Predicting an ICT business process innovation as a digital transformation with machine learning techniques. *Technology Analysis and Strategic Management*, *36*(9), 2271, 2283, doi: 10.1080/09537325.2022.2132927.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427–465. doi: 10.1111/j.1540-6261.1992.tb04398.x.

Gandin, I., & Cozza, C. (2019). Can we predict firms' innovativeness? The identification of innovation performers in an Italian region through a supervised learning approach. *PLoS One*, *14*(6), e0218175. doi: 10.1371/journal.pone.0218175.

Garcia, R., & Martinez, L. (2024). Diversity as value driver in Euro Stoxx 50 companies. *The Journal of Risk Finance*, *25*(3), 471–488. doi:10.1108/JRF-12-2023-0310.

Geng, R., Bose, I., & Chen, X. (2015). Prediction of financial distress: An empirical study of listed Chinese companies using data mining. *European Journal of Operational Research*, *241*(1), 236–247. doi: 10.1016/j.ejor.2014.08.016.

Gentry, R. J., & Shen, W. (2010). The relationship between accounting and market measures of firm financial performance: How strong is it?. *Journal of Managerial Issues*, *22*(4), 514–530. Available from: http://www.jstor.org/stable/25822528

Ghani, B., Hyder, S. I., Yoo, S., & Han, H. (2023). Does employee engagement promote innovation? The facilitators of innovative workplace behavior via mediation and moderation. *Heliyon*, *9*(11), e21817. doi: 10.1016/j.heliyon.2023.e21817.

Giap, C. N., Ha, D. T., Huy, V. Q., Hien, D. T. T., Son, D. T., & Trang, L. M. (2021). Firm performance prediction for macroeconomic diffusion index using machine learning. *International Journal of Advanced Computer Science and Applications*, *12*(6), 87–99. doi: 10.14569/IJACSA.2021.0120612.

Gleißner, W., Günther, T., & Walkshäusl, C. (2022). Financial sustainability: Measurement and empirical evidence. *Journal of Business Economics*, *92*(3), 467–516. doi: 10.1007/s11573-022-01081-0.

Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance*, *32*, 100577. doi: 10.1016/j.jbef.2021.100577.

Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Upper Saddle River, NJ: Pearson Education.

Gupta, A., Rawte, V., & Zaki, M. J. (2023). Predicting firm financial performance from SEC filing changes using automatically generated dictionary. *Computational Economics*, *64*(1), 307–334. doi: 10.1007/s10614-023-10443-x.

Hajek, P., Olej, V., & Myskova, R. (2014). Forecasting corporate financial performance using sentiment in annual reports for stakeholders' decision-making. *Technological and Economic Development of Economy*, *20*(4), 721–738. doi: 10.3846/20294913.2014.979456.

Handoyo, S., Suharman, H., Ghani, E. K., & Soedarsono, S. (2023). A business strategy, operational efficiency, ownership structure, and manufacturing performance: The moderating role of market uncertainty and competition intensity and its implication on open innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, *9*(2), 100039. doi: 10.1016/j.joitmc.2023.100039.

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, *124*, 226–251. doi: 10.1016/j.eswa.2019.01.012.

Heredia, J., Castillo-Vergara, M., Geldes, C., Carbajal Gamarra, F. M., Flores, A., & Heredia, W. (2022). How do digital capabilities affect firm performance? The mediating role of technological capabilities in the new normal. *Journal of Innovation and Knowledge*, *7*(2), 100171. doi: 10.1016/j.jik.2022.100171.

Hezam, Y. A. A., Anthonysamy, L., & Suppiah, S. D. K. (2023). Big data analytics and auditing: A review and synthesis of literature. *Emerging Science Journal*, *7*(2), 629–642. doi: 10.28991/esj-2023-07-02-023.

Hsiao, C. (2014). *Analysis of Panel Data* (3rd ed.). New York: Cambridge University Press.

Jiang, Y., & Jones, S. (2018). Corporate distress prediction in China: A machine learning approach. *Accounting and Finance*, *58*(4), 1063–1109. doi: 10.1111/acfi.12432.

Jin, S. H., & Choi, S. O. (2019). The effect of innovation capability on business performance: A focus on it and business service companies. *Sustainability*, *11*(19), 5246. doi: 10.3390/su11195246. Available from: https://www.mdpi.com/2071-1050/11/19/5246

Jing, W., & Anhang, L. (2023). Design of enterprise financial performance prediction model based on artificial intelligence algorithm. *Soft Computing*, *28*(S2), 607. doi: 10.1007/s00500-023-08516y.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. doi: 10.1126/science.aaa8415.

Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, *74*(7), 2561–2573. doi: 10.1016/j.jpdc.2014.01.003.

Kang, T., Park, D. H., & Han, I. (2018). Beyond the numbers: The effect of 10-K tone on firms' performance predictions using text analytics. *Telematics and Informatics*, *35*(2), 370–381. doi: 10.1016/j.tele.2017.12.014.

Kang, I. G., Kim, N., Loh, W. Y., & Bichelmeyer, B. A. (2021). A machine-learning classification tree model of perceived organizational performance in US federal government health agencies. *Sustainability*, *13*(18), 10329. doi: 10.3390/su131810329.

Khalil, M. L., Aziz, N. A., Long, F., & Zhang, H. (2023). What factors affect firm performance in the hotel industry post-Covid-19 pandemic? Examining the impacts of big data analytics capability, organizational agility and innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, *9*(2), 100081. doi: 10.1016/j.joitmc.2023.100081.

Khan, H., Rehmat, M., Butt, T. H., Farooqi, S., & Asim, J. (2020). Impact of transformational leadership on work performance, burnout and social loafing: A mediation model. *Future Business Journal*, *6*(1), 40. doi: 10.1186/s43093-020-00043-8.

Kim, J., Kim, H., & Geum, Y. (2023). How to succeed in the market? Predicting startup success using a machine learning approach. *Technological Forecasting and Social Change*, *193*, 122614. doi: 10.1016/j.techfore.2023.122614.

Kitchenham, B., Pretorius, R., Budgen, D., Pearl Brereton, O., Turner, M., Niazi, M., & Linkman, S. (2010). Systematic literature reviews in software engineering – a tertiary study. *Information and Software Technology*, *52*(8), 792–805. doi: 10.1016/j.infsof.2010.03.006.

Kossyva, D., Theriou, G., Aggelidis, V., & Sarigiannidis, L. (2023). Outcomes of engagement: A systematic literature review and future research directions. *Heliyon*, *9*(6), e17565. doi: 10.1016/j.heliyon.2023.e17565.

Kureljusic, M., & Karger, E. (2023). Forecasting in financial accounting with artificial intelligence – a systematic literature review and future research agenda. *Journal of Applied Accounting Research*, *25*(1), 81–104. doi: 10.1108/JAAR-06-2022-0146.

Laghari, F., Ahmed, F., & López García, M. L. N. (2023). Cash flow management and its effect on firm performance: Empirical evidence on non-financial firms of China. *PLoS One*, *18*(6), e0287135. doi: 10.1371/journal.pone.0287135.

Lam, M. (2004). Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis. *Decision Support Systems*, *37*(4), 567–581. doi: 10.1016/S0167-9236(03)00088-5.

Lee, J., Kim, G., Park, S., & Jang, D. (2016). Hybrid corporate performance prediction model considering technical capability. *Sustainability (Switzerland)*, *8*(7), 640. doi: 10.3390/su8070640.

Lee, J., Jang, D., & Park, S. (2017). Deep learning-based corporate performance prediction model considering technical capability. *Sustainability (Switzerland)*, *9*(6), 899. doi: 10.3390/su9060899.

Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, *7*(1), 29. doi: 10.3390/risks7010029. Available from: https://www.mdpi.com/2227-9091/7/1/29

Leso, B. H., Cortimiglia, M. N., & Ghezzi, A. (2023). The contribution of organizational culture, structure, and leadership factors in the digital transformation of SMEs: A mixed-methods approach. *Cognition, Technology and Work*, *25*(1), 151–179. doi: 10.1007/s10111-022-00714-2.

Li, H., Sun, J., & Wu, J. (2010). Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. *Expert Systems with Applications*, *37*(8), 5895–5904. doi: 10.1016/j.eswa.2010.02.016.

Li, C., Naz, S., Khan, M. A. S., Kusi, B., & Murad, M. (2019). An empirical investigation on the relationship between a high-performance work system and employee performance: Measuring a mediation model through partial least squares-structural equation modeling. *Psychology Research and Behavior Management*, *12*, 397–416. doi: 10.2147/prbm.S195533.

Li, H., He, L. Y., & Yang, J. J. (2022). Forecasting the medium-term performance of restructured tourism firms with an adaptive integrated predictor. *Tourism Management*, *88*, 104436. doi: 10.1016/j.tourman.2021.104436.

Liu, H. (2021). Research on performance prediction of technological innovation enterprises based on deep learning. *Wireless Communications and Mobile Computing*, *2021*(1), 1682163. doi: 10.1155/2021/1682163.

Liu, H., Wu, S., Zhong, C., & Liu, Y. (2020). The sustainable effect of operational performance on financial benefits: Evidence from Chinese quality awards winners. *Sustainability*, *12*(5), 1966. doi: 10.3390/su12051966. Available from: https://www.mdpi.com/2071-1050/12/5/1966

Loh, W.-Y. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, *1*(1), 14–23. doi: 10.1002/widm.8.

Luong, A., Kumar, N., & Lang, K. R. (2020). Man-machine collaboration in organizational decision-making: Man-machine collaboration in organizational decision-making: An experimental study using loan application evaluations an experimental study using loan application evaluations. *Proceedings of the 2019 pre-ICIS SIGDSA symposium*.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS One*, *13*(3), e0194889. doi: 10.1371/journal.pone.0194889.

Martens, D., Vanhoutte, C., De Winne, S., Baesens, B., Sels, L., & Mues, C. (2011). Identifying financially successful start-up profiles with data mining. *Expert Systems with Applications*, *38*(5), 5794–5800. doi: 10.1016/j.eswa.2010.10.052.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, *90*(10), 60–68. Available from: https://hbr.org/2012/10/big-data-the-management-revolution

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, *98*, 261–276. doi: 10.1016/j.jbusres.2019.01.044.

Mohamad, H. H., Ibrahim, A. H., & Massoud, H. H. (2014). Modelling the financial performance of construction companies using neural network via genetic algorithm. *Canadian Journal of Civil Engineering*, *41*(11), 945–954. doi: 10.1139/cjce-2014-0065.

Moscatelli, M., Parlapiano, F., Narizzano, S., & Viggiano, G. (2020). Corporate default forecasting with machine learning. *Expert Systems with Applications*, *161*, 113567. doi: 10.1016/j.eswa.2020.113567.

Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, *37*(1), 879–910. doi: 10.17705/1cais.03743.

Omar, T., & Zineb, I. (2019). Firm performance: Definition and measurement models. *European Scientific Journal*, *15*(1), 93. doi: 10.19044/esj.2019.v15n1p93.

Onufrey, K., & Bergek, A. (2021). Transformation in a mature industry: The role of business and innovation strategies. *Technovation*, *105*, 102190. doi: 10.1016/j.technovation.2020.102190.

Pap, J., Mako, C., Illessy, M., Kis, N., & Mosavi, A. (2022). Modeling organizational performance with machine learning. *Journal of Open Innovation: Technology, Market, and Complexity*, *8*(4), 177. doi: 10.3390/joitmc8040177.

Park, M. S., Son, H., Hyun, C., & Hwang, H. J. (2021). Explainability of machine learning models for bankruptcy prediction. *IEEE Access*, *9*, 124887–124899. doi: 10.1109/access.2021.3110270.

Porter, M. E. (1985). *Competitive Advantage: Creating and Sustaining Superior Performance*. New York: The Free Press.

Prisma (2020). PRISMA 2020 statement. Available from: https://www.prisma-statement.org/citing-prisma-2020

Qiu, X. Y., Srinivasan, P., & Hu, Y. (2014). Supervised learning models to predict firm performance with annual reports: An empirical study. *Journal of the American Society for Information Science and Technology*, *65*(2), 400–413. doi: 10.1002/asi.22983.

Rai, R., Tiwari, M. K., Ivanov, D., & Dolgui, A. (2021). Machine learning in manufacturing and industry 4.0 applications. *International Journal of Production Research*, *59*(16), 4773–4778. doi: 10.1080/00207543.2021.1956675.

Rajapathirana, R. P. J., & Hui, Y. (2018). Relationship between innovation capability, innovation type, and firm performance. *Journal of Innovation and Knowledge*, *3*(1), 44–55. doi: 10.1016/j.jik.2017.06.002.

Reis, C., Ruivo, P., Oliveira, T., & Faroleiro, P. (2020). Assessing the drivers of machine learning business value. *Journal of Business Research*, *117*, 232–243. doi: 10.1016/j.jbusres.2020.05.053.

Richard, P. J., Devinney, T. M., Yip, G. S., & Johnson, G. (2009). Measuring organizational performance: Towards methodological best practice. *Journal of Management*, *35*(3), 718–804. doi: 10.1177/0149206308330560.

RL, M., & Mishra, A. K. (2022). Measuring financial performance of Indian manufacturing firms: Application of decision tree algorithms. *Measuring Business Excellence*, *26*(3), 288–307. doi: 10.1108/MBE-05-2020-0073.

Robert Baum, J., & Wally, S. (2003). Strategic decision speed and firm performance. *Strategic Management Journal*, *24*(11), 1107–1129. doi: 10.1002/smj.343.

Roumani, Y. F., Nwankpa, J. K., & Tanniru, M. (2020). Predicting firm failure in the software industry. *Artificial Intelligence Review*, *53*(6), 4161–4182. doi: 10.1007/s10462019-09789-2.

Rožman, M., Tominc, P., & Štrukelj, T. (2023). Competitiveness through development of strategic talent management and agile management ecosystems. *Global Journal of Flexible Systems Management*, *24*(3), 373–393. doi: 10.1007/s40171-023-00344-1.

Saha, D., Young, T. M., & Thacker, J. (2023). Predicting firm performance and size using machine learning with a Bayesian perspective. *Machine Learning with Applications*, *11*, 100453. doi: 10.1016/j.mlwa.2023.100453.

Saito, M., Ohsato, T., & Yamanaka, S. (2021). An empirical evaluation of machine learning performance in corporate sales growth prediction. *JSIAM Letters*, *13*, 25–28, doi: 10.14495/jsiaml.13.25.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), 160. doi: 10.1007/s42979-021-00592-x.

Schalck, C., & Yankol-Schalck, M. (2021). Predicting French SME failures: New evidence from machine learning techniques. *Applied Economics*, *53*(51), 5948–5963. doi: 10.1080/00036846.2021.1934389.

Schäper, T., Jung, C., Foege, J. N., Bogers, M. L. A. M., Fainshmidt, S., & Nüesch, S. (2023). The S-shaped relationship between open innovation and financial performance: A longitudinal perspective using a novel text-based measure. *Research Policy*, *52*(6), 104764. doi: 10.1016/j.respol.2023.104764.

Sehgal, S., Mishra, R. K., Deisting, F., & Vashisht, R. (2021). On the determinants and prediction of corporate financial distress in India. *Managerial Finance*, *47*(10), 1428–1447. doi: 10.1108/MF-06-2020-0332.

Sharma, M., Gupta, R., Sehrawat, R., Jain, K., & Dhir, A. (2023). The assessment of factors influencing Big data adoption and firm performance: Evidences from emerging economy. *Enterprise Information Systems*, *17*(12), 2218160. doi: 10.1080/17517575.2023.2218160.

Shie, F. S., Chen, M. Y., & Liu, Y. S. (2012). Prediction of corporate financial distress: An application of the America banking industry. *Neural Computing and Applications*, *21*(7), 1687–1696. doi: 10.1007/s00521-011-0765-5.

Siddiqui, H. U. R., de Abajo, B. S., Diez, I. D., Rustam, F., Raza, A., Atta, S., & Ashraf, I. (2023). Predicting bankruptcy of firms using earnings call data and transfer learning. *PeerJ Computer Science*, *9*, e1134. doi: 10.7717/peerj-cs.1134.

Smith, H., Discetti, R., Bellucci, M., & Acuti, D. (2022). SMEs engagement with the sustainable development goals: a power perspective. *Journal of Business Research*, *149*, 112–122. doi: 10.1016/j.jbusres.2022.05.021.

Song, Y. G., Cao, Q. L., & Zhang, C. (2018). Towards a new approach to predict business performance using machine learning. *Cognitive Systems Research*, *52*, 1004–1012. doi: 10.1016/j.cogsys.2018.09.006.

Sonmez Cakir, F., & Adiguzel, Z. (2020). Analysis of leader effectiveness in organization and knowledge sharing behavior on employees and organization. *SAGE Open*, *10*(1). doi: 10.1177/2158244020914634.

Sun, Z., Sun, L., & Strang, K. (2018). Big data analytics services for enhancing business intelligence. *Journal of Computer Information Systems*, *58*(2), 162–169. doi: 10.1080/08874417.2016.1220239.

Svanberg, J., Ardeshiri, T., Samsten, I., Öhman, P., Neidermeyer, P. E., Rana, T., . . . Danielson, M. (2022). Corporate governance performance ratings with machine learning. *Intelligent Systems in Accounting, Finance and Management*, *29*(1), 50–68. doi: 10.1002/isaf.1505.

Tran, L. T. H., Tu, T. T. K., Nguyen, T. T. H., Nguyen, H. T. L., & Vo, X. V. (2021). Annual report narrative disclosures, information asymmetry and future firm performance: Evidence from Vietnam. *International Journal of Emerging Markets*, *18*(2), 351–375. doi: 10.1108/ijoem-08-2020-0925.

Tran, L. T. H., Tu, T. T. K., Nguyen, T. T. H., Nguyen, H. T. L., & Vo, X. V. (2023). Annual report narrative disclosures, information asymmetry and future firm performance: Evidence from Vietnam. *International Journal of Emerging Markets*, *18*(2), 351–375. doi: 10.1108/IJOEM-08-2020-0925.

Ulutagay, G., Ecer, F., & Nasibov, E. (2015). Performance evaluation of industrial enterprises via fuzzy inference system approach: A case study. *Soft Computing*, *19*(2), 449–458. doi: 10.1007/s00500-014-1263-3.

Unsal, O., & Hassan, M. K. (2023). Employee treatment and firm performance: evidence from topic modelling in lawsuit announcements. *Review of Quantitative Finance and Accounting*, 1–38.

Venkatraman, N., & Ramanujam, V. (1986). Measurement of business performance in strategy research: A comparison of approaches. *The Academy of Management Review*, *11*(4), 801814. doi: 10.2307/258398.

Vukovic, D. B., Spitsina, L., Gribanova, E., Spitsin, V., & Lyzin, I. (2023). Predicting the performance of retail market firms: Regression and machine learning methods. *Mathematics*, *11*(8), 1916. doi: 10.3390/math11081916.

Wang, H. (2021). Research on influencing factors of financial performance of listed companies based on multiple linear regression and fuzzy logic system. *Journal of Intelligent and Fuzzy Systems*, *40*(4), 8549–8561. doi: 10.3233/JIFS-189675.

Wang, A., & Yu, H. (2022). The construction and empirical analysis of the company's financial early warning model based on data mining algorithms. *Journal of Mathematics*, *2022*(1), 3808895. doi: 10.1155/2022/3808895.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). MA: MIT Press.

Yoo, H. S., Jung, Y. L., & Jun, S. P. (2023). Prediction of SMEs' R&D performances by machine learning for project selection. *Scientific Reports*, *13*(1), 7598. doi: 10.1038/s41598-023-34684-w.

Zaini, B. J., & Mahmuddin, M. (2019). Classifying firms' performance using data mining approaches. *International Journal of Supply Chain Management*, *8*(1), 690–696. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85063266971&partnerID=40&md5=413dd3f1d978683f5cfc4d7cf15b8842

Zhang, Z., Xiao, Y., & Niu, H. (2022). DEA and machine learning for performance prediction. *Mathematics*, *10*(10), 1776. doi: 10.3390/math10101776.

Zhang, R., Tian, Z., McCarthy, K. J., Wang, X., & Zhang, K. (2023). Application of machine learning techniques to predict entrepreneurial firm valuation. *Journal of Forecasting*, *42*(2), 402–417. doi: 10.1002/for.2912.

Zhu, N., Zhu, C., & Emrouznejad, A. (2021). A combined machine learning algorithms and DEA method for measuring and predicting the efficiency of Chinese manufacturing listed companies. *Journal of Management Science and Engineering*, *6*(4), 435–448. doi: 10.1016/j.jmse.2020.10.001.

Appendix

Table A1. Attributes frequently used in predicting firm performance

Attributes	Attributes domain	Frequency	Reference
Financial performance	This attribute can include metrics such as revenue growth, profit margins, and financial ratios (e.g. liquidity, solvency)	46	Giap *et al.* (2021), Zhang, Xiao, and Niu (2022), Vukovic *et al.* (2023), Cozgarea *et al.* (2023), Saito *et al.* (2021), Song, Cao, and Zhang (2018), Boubaker, Le, Ngo, and Manita (2023), Martens *et al.* (2011), Alanis, Chava, and Shah (2023), Yoo, Jung, and Jun (2023), Colak, Fu, and Hasan (2020), Kim, Kim, and Geum (2023), Jiang and Jones (2018), Zhu, Zhu, and Emrouznejad (2021), Dong, Wang, and Cao (2022), Schalck and Yankol-Schalck (2021), Creamer and Freund (2010), Geng, Bose, and Chen (2015), Awoin, Appiahene, Gyasi, and Sabtiwu (2020), Wang and Yu (2022), Ulutagay, Ecer, and Nasibov (2015), Li, He, and Yang (2022), Lee, Jang, and Park (2017), Gupta *et al.* (2023), Hajek *et al.* (2014), Wang (2021), Unsal and Hassan (2023), Lam (2004), Qiu, Srinivasan, and Hu (2014), Tran, Tu, Nguyen, Nguyen, and Vo (2021), Ben Jabeur, Stef, and Carmona (2023), Che, Zhu, and Li (2020), Bahrami *et al.* (2023), Siddiqui *et al.* (2023), Gandin and Cozza (2019), Shie, Chen, and Liu (2012), Elamir (2021), Zaini and Mahmuddin (2019), Kang, Park, and Han (2018), Mohamad, Ibrahim, and Massoud (2014), Delen, Kuzey, and Uyar (2013), Jing and Anhang (2023), Liu (2021), Lee, Kim, Park, and Jang (2016), Bargagli-Stoffi *et al.* (2021), Chen and Du (2009), Li, Sun, and Wu (2010), Moscatelli, Parlapiano, Narizzano, and Viggiano (2020), Pap *et al.* (2022), RL and Mishra (2022), Saha *et al.* (2023), Sehgal, Mishra, Deisting, and Vashisht (2021)
Market position	Evaluate factors like market share, customer base, and competitive analysis	23	Zhang *et al.* (2022), Vukovic *et al.* (2023), Martens *et al.* (2011), Alanis *et al.* (2023), Yoo *et al.* (2023), Jiang and Jones (2018), Creamer and Freund (2010), Ulutagay *et al.* (2015), Wang (2021), Lam (2004), Tran, Tu, Nguyen, Nguyen, and Vo (2023), Ben Jabeur *et al.* (2023), Che *et al.* (2020), Bahrami *et al.* (2023), Siddiqui *et al.* (2023), Shie *et al.* (2012), Elamir (2021), Kang *et al.* (2018), Hajek *et al.* (2014), Mohamad *et al.* (2014), Delen *et al.* (2013), Lee *et al.* (2016), Li *et al.* (2010)

(*continued*)

Table A1. Continued

Attributes	Attributes domain	Frequency	Reference
Operational efficiency	Assess factors like production efficiency, supply chain effectiveness, and cost management	22	Giap *et al.* (2021), Zhang *et al.* (2022), Boubaker *et al.* (2023), Kim *et al.* (2023), Zhu *et al.* (2021), Dong *et al.* (2022), Creamer and Freund (2010), Ulutagay *et al.* (2015), Li *et al.* (2022), Unsal and Hassan (2023), Tran *et al.* (2023), Ben Jabeur *et al.* (2023), Che *et al.* (2020), Siddiqui *et al.* (2023), Gandin and Cozza (2019), Shie *et al.* (2012), Kang *et al.* (2018), Hajek *et al.* (2014) Mohamad *et al.* (2014), Delen *et al.* (2013), Lee *et al.* (2016), Li *et al.* (2010)
Innovation capability	Consider research and development spending, patents, and innovation awards	12	Kang, Kim, Loh, and Bichelmeyer (2021), Martens *et al.* (2011), Zhang, Tian, McCarthy, Wang, and Zhang (2023), Eom, Woo, and Chun (2022), Yoo *et al.* (2023), Kim *et al.* (2023), Lee *et al.* (2016, 2017), Che *et al.* (2020), Gandin and Cozza (2019), Jing and Anhang (2023), Liu (2021), Pap *et al.* (2022)
Leadership quality	Include attributes related to the leadership team, such as experience, qualifications, and track record	10	Giap *et al.* (2021), Svanberg *et al.* (2022), Cozgarea *et al.* (2023), Kang *et al.* (2021), Zhang *et al.* (2023), Kim *et al.* (2023), Jiang and Jones (2018), Creamer and Freund (2010), Che *et al.* (2020), Shie *et al.* (2012)
Employee engagement	Measure factors like employee satisfaction, turnover rate, and skill development	12	Zhang *et al.* (2022), Svanberg *et al.* (2022), Cozgarea *et al.* (2023), Boubaker *et al.* (2023), Kang *et al.* (2021), Martens *et al.* (2011), Kim *et al.* (2023), Jiang and Jones (2018), Creamer and Freund (2010), Ulutagay *et al.* (2015), Che *et al.* (2020), Pap *et al.* (2022), Shie *et al.* (2012)

Source(s): Table by authors

Table A2. Performance prediction methods

Features	Method	Count	Reference
Financial performance	Classification	43	Alanis *et al.* (2023), Bahrami *et al.* (2023), Ben Jabeur *et al.* (2023), Boubaker *et al.* (2023), Che *et al.* (2020), Colak *et al.* (2020), Cozgarea *et al.* (2023), Creamer and Freund (2010), Delen *et al.* (2013), Elamir (2021), Gandin and Cozza (2019), Geng *et al.* (2015), Giap *et al.* (2021), Gupta *et al.* (2023), Hajek *et al.* (2014), Jiang and Jones (2018), Jing and Anhang (2023), Kang *et al.* (2018), Kim *et al.* (2023), Lam (2004), Lee *et al.* (2017), Lee *et al.* (2016), Li *et al.* (2022), Li *et al.* (2010), Liu (2021), Martens *et al.* (2011), Mohamad *et al.* (2014), Qiu *et al.* (2014), Saito *et al.* (2021), Schalck and Yankol-Schalck (2021), Shie *et al.* (2012), Siddiqui *et al.* (2023), Song *et al.* (2018), Tran *et al.* (2023), Unsal and Hassan (2023), Vukovic *et al.* (2023), Wang and Yu (2022), Wang (2021), Yoo *et al.* (2023), Zaini and Mahmuddin (2019), Zhang *et al.* (2022), Zhu *et al.* (2021)
	Regression	22	Alanis *et al.* (2023), Bahrami *et al.* (2023), Ben Jabeur *et al.* (2023), Boubaker *et al.* (2023), Che *et al.* (2020), Colak *et al.* (2020), Cozgarea *et al.* (2023), Delen *et al.* (2013), Elamir (2021), Gupta *et al.* (2023), Jiang and Jones (2018), Jing and Anhang (2023), Kang *et al.* (2018), Lam (2004), Lee *et al.* (2016), Qiu *et al.* (2014), Schalck and Yankol-Schalck (2021), Shie *et al.* (2012), Siddiqui *et al.* (2023), Unsal and Hassan (2023), Wang (2021), Zaini and Mahmuddin (2019)
	Clustering	32	Alanis *et al.* (2023), Awoin *et al.* (2020), Bahrami *et al.* (2023), Boubaker *et al.* (2023), Colak *et al.* (2020), Cozgarea *et al.* (2023), Creamer and Freund (2010), Delen *et al.* (2013), Dong *et al.* (2022), Elamir (2021), Gandin and Cozza (2019), Geng *et al.* (2015), Giap *et al.* (2021), Hajek *et al.* (2014), Jing and Anhang (2023), Kang *et al.* (2018), Kim *et al.* (2023), Lam (2004), Li *et al.* (2022), Martens *et al.* (2011), Qiu *et al.* (2014), Saito *et al.* (2021), Shie *et al.* (2012), Siddiqui *et al.* (2023), Song *et al.* (2018), Ulutagay *et al.* (2015), Unsal and Hassan (2023), Vukovic *et al.* (2023), Wang and Yu (2022), Yoo *et al.* (2023), Zaini and Mahmuddin (2019), Zhang *et al.* (2022)
Market position	Classification	24	Alanis *et al.* (2023), Bahrami *et al.* (2023), Ben Jabeur *et al.* (2023), Che *et al.* (2020), Creamer and Freund (2010), Delen *et al.* (2013), Elamir (2021), Giap *et al.* (2021), Hajek *et al.* (2014), Jiang and Jones (2018), Kang *et al.* (2018, 2021), Kim *et al.* (2023), Lam (2004), Lee *et al.* (2016), Martens *et al.* (2011), Mohamad *et al.* (2014), Shie *et al.* (2012), Siddiqui *et al.* (2023), Tran *et al.* (2023), Vukovic *et al.* (2023), Wang (2021), Yoo *et al.* (2023), Zhang *et al.* (2022)
	Regression	13	Alanis *et al.* (2023), Bahrami *et al.* (2023), Ben Jabeur *et al.* (2023), Che *et al.* (2020), Delen *et al.* (2013), Elamir (2021), Jiang and Jones (2018), Kang *et al.* (2018), Lee *et al.* (2016), Shie *et al.* (2012), Siddiqui *et al.* (2023), Wang (2021), Zhang *et al.* (2023)
	Clustering	17	Alanis *et al.* (2023), Bahrami *et al.* (2023), Creamer and Freund (2010), Delen *et al.* (2013), Elamir (2021), Giap *et al.* (2021), Hajek *et al.* (2014), Kang *et al.* (2018), Kim *et al.* (2023), Lam (2004), Qiu *et al.* (2014), Shie *et al.* (2012), Siddiqui *et al.* (2023), Ulutagay *et al.* (2015), Vukovic *et al.* (2023), Yoo *et al.* (2023), Zhang *et al.* (2022)

(*continued*)

Table A2. Continued

Features	Method	Count	Reference
Operational efficiency	Classification	21	Ben Jabeur *et al.* (2023), Boubaker *et al.* (2023), Che *et al.* (2020), Creamer and Freund (2010), Delen *et al.* (2013), Gandin and Cozza (2019), Giap *et al.* (2021), Hajek *et al.* (2014), Kang *et al.* (2018), Kim *et al.* (2023), Lee *et al.* (2016), Li *et al.* (2022), Martens *et al.* (2011), Mohamad *et al.* (2014), Shie *et al.* (2012), Siddiqui *et al.* (2023), Tran *et al.* (2021, 2023), Unsal and Hassan (2023), Zhang *et al.* (2022), Zhu *et al.* (2021)
	Regression	10	Ben Jabeur *et al.* (2023), Boubaker *et al.* (2023), Che *et al.* (2020), Delen *et al.* (2013), Kang *et al.* (2018), Lam (2004), Lee *et al.* (2016), Shie *et al.* (2012), Siddiqui *et al.* (2023), Unsal and Hassan (2023)
	Clustering	16	Boubaker *et al.* (2023), Creamer and Freund (2010), Delen *et al.* (2013), Dong *et al.* (2022), Gandin and Cozza (2019), Giap *et al.* (2021), Hajek *et al.* (2014), Kang *et al.* (2018), Kim *et al.* (2023), Lam (2004), Li *et al.* (2022), Shie *et al.* (2012), Siddiqui *et al.* (2023), Song *et al.* (2018), Ulutagay *et al.* (2015), Unsal and Hassan (2023), Zhang *et al.* (2022)
Innovation capability	Classification	10	Che *et al.* (2020), Gandin and Cozza (2019), Jing and Anhang (2023), Kim *et al.* (2023), Lee *et al.* (2016, 2017), Liu (2021), Martens *et al.* (2011), Yoo *et al.* (2023), Zhang *et al.* (2023)
	Regression	3	Che *et al.* (2020), Jing and Anhang (2023), Lee *et al.* (2016)
	Clustering	6	Eom *et al.* (2022), Gandin and Cozza (2019), Jing and Anhang (2023), Kang *et al.* (2021), Kim *et al.* (2023), Yoo *et al.* (2023)
Leadership quality	Classification	10	Che *et al.* (2020), Cozgarea *et al.* (2023), Creamer and Freund (2010), Giap *et al.* (2021), Jiang and Jones (2018), Kim *et al.* (2023), Martens *et al.* (2011), Shie *et al.* (2012), Svanberg *et al.* (2022), Zhang *et al.* (2023)
	Regression	4	Che *et al.* (2020), Cozgarea *et al.* (2023), Jiang and Jones (2018), Shie *et al.* (2012)
	Clustering	7	Cozgarea *et al.* (2023), Creamer and Freund (2010), Giap *et al.* (2021), Kang *et al.* (2021), Kim *et al.* (2023), Shie *et al.* (2012), Svanberg *et al.* (2022)
Employee engagement	Classification	10	Boubaker *et al.* (2023), Che *et al.* (2020), Cozgarea *et al.* (2023), Creamer and Freund (2010), Jiang and Jones (2018), Kim *et al.* (2023), Martens *et al.* (2011), Shie *et al.* (2012), Svanberg *et al.* (2022), Zhang *et al.* (2022)
	Regression	5	Boubaker *et al.* (2023), Che *et al.* (2020), Cozgarea *et al.* (2023), Jiang and Jones (2018), Shie *et al.* (2012)
	Clustering	9	Boubaker *et al.* (2023), Cozgarea *et al.* (2023), Creamer and Freund (2010), Kang *et al.* (2021), Kim *et al.* (2023), Shie *et al.* (2012), Svanberg *et al.* (2022), Ulutagay *et al.* (2015), Zhang *et al.* (2022)

Source(s): Table by authors

Table A3. ML algorithms used in firm performance prediction

Algorithms	Reference	Count
Neural network	Alanis *et al.* (2023), Awoin *et al.* (2020), Bahrami *et al.* (2023), Ben Jabeur *et al.* (2023), Boubaker *et al.* (2023), Elamir (2021), Geng *et al.* (2015), Hajek *et al.* (2014), Jing and Anhang (2023), Lam (2004), Lee *et al.* (2017), Liu (2021), Mohamad *et al.* (2014), Qiu *et al.* (2014), Roumani *et al.* (2020), Shie *et al.* (2012), Siddiqui *et al.* (2023), Song *et al.* (2018), Svanberg *et al.* (2022), Unsal and Hassan (2023), Vukovic *et al.* (2023), Wang and Yu (2022), Yoo *et al.* (2023), Zaini and Mahmuddin (2019), Zhang *et al.* (2023), Zhang *et al.* (2022), Zhu *et al.* (2021)	27
Logistic regression	Alanis *et al.* (2023), Bahrami *et al.* (2023), Ben Jabeur *et al.* (2023), Boubaker *et al.* (2023), Che *et al.* (2020), Colak *et al.* (2020), Cozgarea *et al.* (2023), Creamer and Freund (2010), Delen *et al.* (2013), Elamir (2021), Gupta *et al.* (2023), Jiang and Jones (2018), Jing and Anhang (2023), Kang *et al.* (2018), Kim *et al.* (2023), Lee *et al.* (2016), Li *et al.* (2022), Qiu *et al.* (2014), Schalck and Yankol-Schalck (2021), Shie *et al.* (2012), Siddiqui *et al.* (2023), Svanberg *et al.* (2022), Unsal and Hassan (2023), Wang (2021), Zaini and Mahmuddin (2019)	26
Naïve bayes	Bahrami *et al.* (2023), Che *et al.* (2020), Colak *et al.* (2020), Giap *et al.* (2021), Hajek *et al.* (2014), Jing and Anhang (2023), Kang *et al.* (2018), Kim *et al.* (2023), Lam (2004), Lee *et al.* (2016), Li *et al.* (2022), Shie *et al.* (2012), Svanberg *et al.* (2022), Tran *et al.* (2021), Unsal and Hassan (2023), Zaini and Mahmuddin (2019)	16
Support vector	Awoin *et al.* (2020), Bahrami *et al.* (2023), Ben Jabeur *et al.* (2023), Boubaker *et al.* (2023), Che *et al.* (2020), Gandin and Cozza (2019), Geng *et al.* (2015), Giap *et al.* (2021), Gupta *et al.* (2023), Hajek *et al.* (2014), Jing and Anhang (2023), Kim *et al.* (2023), Lam (2004), Lee *et al.* (2016), Li *et al.* (2022), Liu (2021), Martens *et al.* (2011), Qiu *et al.* (2014), Saito *et al.* (2021), Siddiqui *et al.* (2023), Song *et al.* (2018), Svanberg *et al.* (2022), Yoo *et al.* (2023), Zaini and Mahmuddin (2019), Zhu *et al.* (2021)	25
K-nearest	Bahrami *et al.* (2023), Cozgarea *et al.* (2023), Delen *et al.* (2013), Elamir (2021), Gandin and Cozza (2019), Hajek *et al.* (2014), Jing and Anhang (2023), Kang *et al.* (2018), Kim *et al.* (2023), Lam (2004), Li *et al.* (2022), Liu (2021), Qiu *et al.* (2014), Shie *et al.* (2012), Siddiqui *et al.* (2023), Unsal and Hassan (2023), Yoo *et al.* (2023)	17
Linear regression	Alanis *et al.* (2023), Bahrami *et al.* (2023), Ben Jabeur *et al.* (2023), Boubaker *et al.* (2023), Che *et al.* (2020), Colak *et al.* (2020), Cozgarea *et al.* (2023), Delen *et al.* (2013), Dong *et al.* (2022), Elamir (2021), Gupta *et al.* (2023), Jiang and Jones (2018), Jing and Anhang (2023), Kang *et al.* (2018), Lee *et al.* (2016), Li *et al.* (2010), Qiu *et al.* (2014), Schalck and Yankol-Schalck (2021), Shie *et al.* (2012), Unsal and Hassan (2023), Vukovic *et al.* (2023), Wang (2021), Zaini and Mahmuddin (2019)	22
Decision trees	Alanis *et al.* (2023), Awoin *et al.* (2020), Bahrami *et al.* (2023), Colak *et al.* (2020), Cozgarea *et al.* (2023), Creamer and Freund (2010), Delen *et al.* (2013), Elamir (2021), Gandin and Cozza (2019), Geng *et al.* (2015), Giap *et al.* (2021), Hajek *et al.* (2014), Jing and Anhang (2023), Kang *et al.* (2021), Kim *et al.* (2023), Lam (2004), Li *et al.* (2022), Li *et al.* (2010), Saito *et al.* (2021), Shie *et al.* (2012), Siddiqui *et al.* (2023), Ulutagay *et al.* (2015), Wang and Yu (2022), Yoo *et al.* (2023), Zaini and Mahmuddin (2019), Zhang *et al.* (2022)	26
Random forest	Alanis *et al.* (2023), Bahrami *et al.* (2023), Boubaker *et al.* (2023), Colak *et al.* (2020), Cozgarea *et al.* (2023), Creamer and Freund (2010), Dong *et al.* (2022), Eom *et al.* (2022), Gandin and Cozza (2019), Giap *et al.* (2021), Hajek *et al.* (2014), Jing and Anhang (2023), Kang *et al.* (2018), Kim *et al.* (2023), Lam (2004), Qiu *et al.* (2014), Saito *et al.* (2021), Shie *et al.* (2012), Siddiqui *et al.* (2023), Song *et al.* (2018), Svanberg *et al.* (2022), Unsal and Hassan (2023), Vukovic *et al.* (2023), Yoo *et al.* (2023)	24

Source(s): Table by authors

Corresponding author
Yaseen Hezam can be contacted at: yaseen.hezam@postgrad.otago.ac.nz