Humanities & Social Sciences Communications

Reference #2

ARTICLE

https://doi.org/10.1057/s41599-025-04361-6 OPEN

Integrating machine learning into business and management in the age of artificial intelligence

Aglaya Batz [1✉], David F. D'Croz-Barón[2], Carlos Jesús Vega Pérez[3] & Carlos A. Ojeda-Sanchez[1]

Machine learning, with its capacity to leverage computational techniques for experiential learning, has profoundly influenced various disciplines, including business and management. Despite its contributions to the progress of these fields and the advent of artificial intelligence presenting new challenges, there remains ambiguity regarding the specific areas of significant advancement and those with potential for further development. This study addresses three central questions: (1) How is the intellectual landscape of machine learning in business and management research organized and structured? (2) What are the primary applications of machine learning in business administration? And (3) What strategic considerations should companies adopt to effectively leverage machine learning in their business applications? By means of co-occurrence analysis of over 9399 peer-reviewed documents retrieved from Scopus discussing machine learning in business and management, we identified fifteen clusters within the literature. This classification serves as a starting point for firms looking to integrate ML into their routines across fifteen distinct topics. Although some firms have appropriated ML, the upsurge of artificial intelligence presents new challenges, including the digital divide, infrastructure and acquisition dilemmas, security concerns especially with outsourced services, and cost-effectiveness in algorithm selection and practical applications.

[1] Universidad del Rosario, Bogotá, Colombia. [2] D&N Business Intelligence & Consulting S.A.S., Bogotá, Colombia. [3] Uptime Analytics S.A.S, Bogotá, Colombia. ✉email: aglaya.batz@urosario.edu.co

Introduction

Machine Learning (ML), a facet of artificial intelligence (AI), utilizes computational methods to boost system performance by learning from experience, underpinned by mathematics, statistics, and computer science (Zhou 2021). Arthur Samuel of IBM coined the term in 1959, defining it as the study of enabling computers to learn without explicit programming. In the past, solving computer-based problems required intricate algorithm development, often posing significant challenges. In contrast, ML empowers programmers to craft problem-specific learning algorithms, enabling computers to autonomously acquire knowledge. Here, the machine leverages data and the algorithm's learning framework to construct effective problem-solving models.

Throughout history, the concept of enabling computers to learn has taken on various names, including pattern recognition, data mining, knowledge discovery, predictive analytics, statistical modelling, adaptive systems, data science, and self-organizing systems. In this work, we broadly categorize these concepts as ML. Scholars have made significant contributions to the field of machine learning (ML) in business administration. ML applications in this area are well-established and have evolved across various domains. For instance, Wenzel et al. (2019) explored ML's role in supply chain management (SCM), while Paschek et al. (2017) highlighted its adoption in business process management (BPM) and digital transformation within enterprises. Pallathadka et al. (2023) discussed ML's impact on agriculture, healthcare, and management, among others. In risk management, ML has been extensively examined, as evidenced by Aziz and Dowling (2019) and Bertolini et al. (2021). The literature also covers ML's integration in marketing (Ngai and Wu, 2022), human resources management (Menon et al. 2024), and Industry 4.0 processes (Bertolini et al. 2021).

While studies highlight practical applications, they also acknowledge limitations. For example, Wenzel et al. (2019) provided an extensive overview of ML methods in SCM but noted gaps in the business, data understanding, and evaluation phases. Kraus et al. (2020) pointed out limitations in deep learning, such as extensive tuning requirements and challenges in modelling behaviours. Non-technical limitations in risk management, including data access barriers and lack of transparency, were emphasized by Aziz and Dowling (2019).

Although these studies advance our understanding of ML in business processes, there is still a need for a comprehensive taxonomy or overview of ML applications across all fields of business administration. The literature has provided in-depth insights into specific areas like marketing and supply chain, but a broader study visualizing ML applications in various business domains is lacking. Hence, this study uniquely contributes to the literature by providing a comprehensive taxonomy of machine learning applications across all fields of business administration, an area previously lacking holistic analysis.

Moreover, the contemporary surge in ML applications, exemplified by the exponential growth of the user base witnessed by ChatGPT since its launch on November 20, 2022 (Hu 2023), has sparked inquiries encompassing ethical considerations, usability concerns, and the anticipated impact of ML and AI across various industries. Notably, one significant enquiry revolves around the transformation of existing business models and process optimization within the domain of business and management. Yet, the fusion of artificial intelligence and algorithm development for process optimization, enhanced decision-making, and streamlined data management within companies is not new and, as pointed out before has been in progress for years.

Hence, our research aims to holistically examine the multifaceted applications of ML in the business and management domains, shedding light on emerging thematic areas at the forefront of this dynamic landscape. To achieve this, we analyse scientific and academic publications spanning the past decades in this field. Employing co-word analysis, a bibliometric algorithm with a proven track record in unveiling salient topics and themes (Callon et al. 1991; van Eck and Waltman, 2007), we aim to construct a comprehensive map of the subject. Our curated database provides a valuable foundation for addressing three pivotal research questions:

1. How is the intellectual landscape of machine learning in business and management research organized and structured?
2. What are the primary applications of machine learning in business administration?
3. What strategic considerations should companies adopt to effectively leverage machine learning in their business applications?

The first question seeks to map out and categorize the various research themes and areas where machine learning is applied within the fields of business and management. It aims to provide a comprehensive overview of the current state of academic research in this domain. Consequently, the second question aims to identify and describe the key areas where machine learning techniques are being effectively utilized in business administration, highlighting practical implementations and their impact on business processes. Lastly, the third question seeks to explore the critical factors and best practices that companies need to consider to successfully integrate machine learning technologies into their business operations. It covers aspects such as technological infrastructure, skill development, data management, ethical considerations, and organizational change.

To address the research questions outlined above, we first retrieved scholarly articles and papers related to machine learning in Business and Management from the SCOPUS database—which encompasses a comprehensive collection of highly ranked, peer-reviewed journals—as detailed in the "Methods" section. We then applied co-word analysis to construct a network based on the authors' keywords extracted from these publications, enabling the identification of clusters formed by the co-occurrence of keyword pairs within individual articles. As elaborated in the "Results" section, this analysis revealed 15 distinct clusters, each representing a unique domain where machine learning techniques are applied. These clusters were plotted on a strategic diagram according to their centrality and density metrics, facilitating the detection of mainstream areas, emerging themes, and domains declining in relevance. Finally, to present a comprehensive taxonomy of machine learning applications in the field of business and management, we consolidated these clusters into five major macro-topics: finance, customer relationship management, decision-making support, innovation and public policy, and data management and sustainability. This taxonomy provides a valuable foundation for future research endeavours and practical implementations, reflecting the current state of how machine learning and artificial intelligence are reshaping business landscapes.

Methods

To answer our research questions effectively, we utilize co-word analysis to examine the interplay between ML and the fields of Business and Management. This analysis relies on academic and scientific papers found in citation databases, assuming that the keywords selected by the authors accurately represent the essence of their research (Callon et al. 1991). In Fig. 1 we provide a

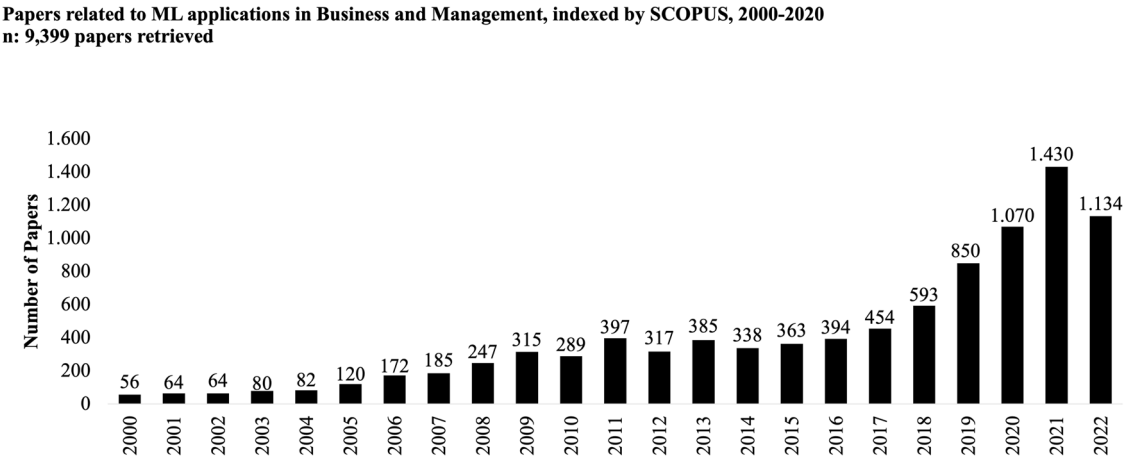


Fig. 1 Number of ML and AI publications in business and management, per year, 2000–mid August 2022. Source: Author's own elaboration.

detailed overview of our methodological approach for conducting co-word analysis.

Data acquisition. Our approach begins with the systematic retrieval of scientific literature on ML in the context of Business and Management. We use the SCOPUS database, which indexes high-quality and highly-ranked journals, and most papers indexed by the Web of Science databases and collections are already indexed by SCOPUS (Singh et al. 2021). As of August 2022, the literature indexed by SCOPUS, published from 2000 onwards, consists of 6940 articles and 2459 conference papers written in English, for a total of 9399 peer-reviewed documents found by the search query (1), and limited to business and management (4834), economics and finance (2422), decision sciences (3074), and social sciences (3158) outlets. Data retrieved for each paper was stored in CSV files, including author names, keywords and affiliations, titles and abstracts, number of citations, and references.

```
(TITLE("machine learning" OR "reinfor* learning" OR "data mining" OR
"*supervised learning" OR clustering OR "dimension* reduction" OR
"*structured prediction" OR "anomaly detection" OR "artificial neural
network" OR "deep learning") OR AUTHKEY("machine learning" OR "reinfor*
learning" OR "data mining" OR "*supervised learning" OR clustering OR
"dimension* reduction" OR "*structured prediction" OR "anomaly
detection" OR "artificial neural network" OR "deep learning"))
```
(1)

The SCOPUS search query (1) focuses on ML and eleven related concepts and techniques, namely reinforced learning, supervised and unsupervised learning, dimensionality reduction, clustering, anomaly detection, structured and unstructured prediction, deep learning (DL), artificial neural networks, and data mining. No additional restrictions were set at this point, domain-wise.

The work of Kemper et al. (1983) is one of the earliest use cases of ML in business and management. Hierarchical clustering (HC) and multidimensional scaling (MDS), two ML approaches, were used in this study to examine tourism as a cultural domain (Kemper et al. 1983). From then on, different ML applications in tourism have been studied by academia. Conversely, pioneer ML applications in finance can be found in the early 1990s (Kaastra and Boyd, 1996). Figure 2 depicts the distribution of retrieved articles over time and demonstrates that 5077 out of 9399 manuscripts, or more than half, have been published since 2018. The growing trend shown over the past 4 years is anticipated to continue.

Keyword pool and network construction. From the literature gathered using the search query (1), we initially extracted 22,103 author keywords, following which we refined the dataset by excluding keywords that appeared in fewer than four papers, aligning with Zhao et al.'s recommendations (2018, p. 3). Using the VOSViewer Software (Waltman et al. 2010) and its *Thesaurus File* functionality, which is used to replace, merge, or remove author keywords, we performed the following preprocessing tasks over the keyword pool: (1) singulars and plurals were standardized (e.g. sensor → sensors), (2) British spellings were merged with their American equivalents (e.g. consumer behaviour → Consumer Behaviour), (3) abbreviations were expanded (e.g. DNN → deep neural network, CART → classification and regression trees), (4) research-related terms were removed (e.g. questionnaire, systematic review), (5) as well as geographical references, (6) the terms employed in the search query were removed, and (7) hyphenated expressions were normalized (e.g. energy-efficient → energy Efficient), resulting in 1162 keywords overall, with 1156 co-occurring in the retrieved papers. A co-occurrence matrix was then constructed, and these co-occurrences were normalized using Noack's LinLog method (2007).

Network visualization. From the pre-processed keyword pool in its normalized co-occurrence matrix form, the co-word network was elaborated with the VOSviewer software. This network is

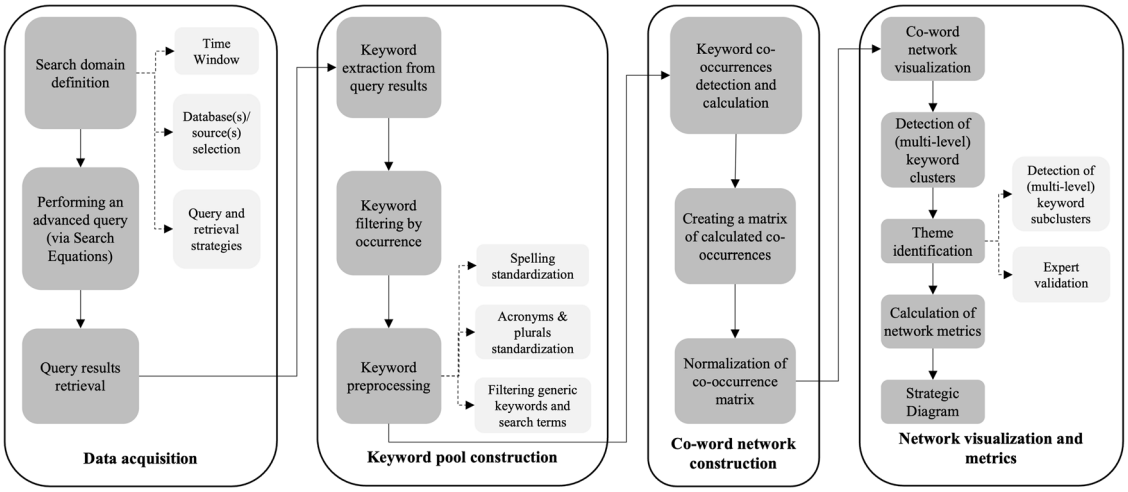


Fig. 2 Methodology for co-word analysis of scientific publications containing author keywords. Source: Author's own elaboration.

depicted as an undirected graph (co-word graph), where the *nodes* represent author keywords, and the *edges* their co-occurrences. Therefore, bigger *nodes* depict higher keyword appearances in retrieved papers (that is, keyword frequencies), and thicker *nodes* depict more frequent keyword co-occurrences (that is, *link strengths*).

As the *Visualization of Similarities* (van Eck and Waltman, 2007) approach built into VOSViewer allows for a simultaneous depiction of co-word networks (VOS mapping) and community detection of keyword groups (VOS clustering), fifteen groups, or clusters of ML in business and management, were detected using the *Modularity* variant of VOS mapping based on the work of Newman and Girvan (2004), with *Attraction* and *Repulsion* parameters set to 1 and −1, respectively. The *Attraction* and *Repulsion* parameters are used by VOS mapping to find the optimal location of each keyword in the co-word graph based on their frequencies and co-occurrences (Waltman et al. 2010), and are mostly used to improve the readability of the co-word graph. Furthermore, the clustering *Resolution* is involved in this stage as a third parameter, where higher values would reveal more, smaller clusters at the expense of interpretation. The default *Resolution* value of 1.0 is sufficient in most cases to reveal coherent clusters. In ours, however, a higher value of 1.5 was investigated, and after evaluating the separation and semantic coherence of revealed clusters, by reviewing the keywords with the most co-occurrences in each, we opted for this value, as it yielded better clusters than those found with *Resolution* 1.0.

Cluster metrics and trends. From the co-word graph and keyword clusters revealed by VOSviewer, two metrics of interest can be obtained, namely *Cluster Centrality* and *Cluster Density*. VOSViewer, however, lacks the required functionality to calculate these metrics within the software. Hence, we developed a computer script for extracting the required data from VOSViewer outputs to calculate both *centralities*, each defined as follows:

- *Cluster centrality* for a given cluster L -C_L-, as modelled by Eq. (2), measures how influential a given cluster is over the research field based on co-occurrences with keywords located elsewhere (Callon et al. 1991).

$$C_L = \sum_{i \in L} \sum_{j \notin L} w_{ij} e_{ij}$$
$$e_{ij} = \begin{cases} 1, \text{if } w_{ij} > 0 \\ 0, \text{otherwise} \end{cases} \quad (2)$$

For each co-occurrence of keywords located in L with those from other clusters (condition modelled by e_{ij}), the corresponding link strength -w_{ij}- is incorporated into the C_L summation.

- *Cluster Density* for a given cluster L -D_L-, as described by Eq. (3), is a ratio of the observed keyword co-occurrences within, over the maximum theoretical co-occurrences allowed for it (Callon et al. 1991). Therefore, the more co-occurrences between the keywords conforming L, the higher the density, and the more developed is the theme described by it.

$$D_L = \frac{2E}{N(N-1)} \quad (3)$$

The cluster metrics modelled above describe the maturity and cohesion of themes identified in the co-word network via a *Strategic Diagram* (Callon et al. 1991). In this *diagram*, keyword clusters are arranged over a two-dimensional plane, where the horizontal axis (X) depicts the cluster densities and the vertical (Y) the cluster densities. Clusters are arranged left to right based on their centralities, where the right-most clusters are the most central ones, and top to bottom after their densities, where the top-most ones are the densest. From the resulting X–Y placement of clusters over the *diagram* and the average cluster centrality and density, four strategic quadrants emerge, each revealing the current and future trends for ML in business themes: The major themes are placed at the top right quadrant (quadrant I), whereas the complementary and peripheral themes are located on the left side of the diagram (quadrants II and III). Lastly, emerging and developing themes are found at the bottom quadrants of the diagram (quadrants III and IV), highlighting them as themes with research potential.

Results

The co-occurrence map of keywords, illustrated in Fig. 3, shows keywords as circles (nodes) and their correlations as curved lines (arcs). A correlation between keywords is indicated when a pair of keywords appears together in at least one retrieved paper (co-occurrence). The literature is organized into 15 distinct clusters: (1) financial markets assessment and forecasting; (2) time-series forecasting; (3) complex network analysis; (4) marketing and sales optimization; (5) digital transformation and sustainability; (6) strategic decision making; (7) education and skills development; (8) opinion mining and Electronic Word of Mouth (eWOM)

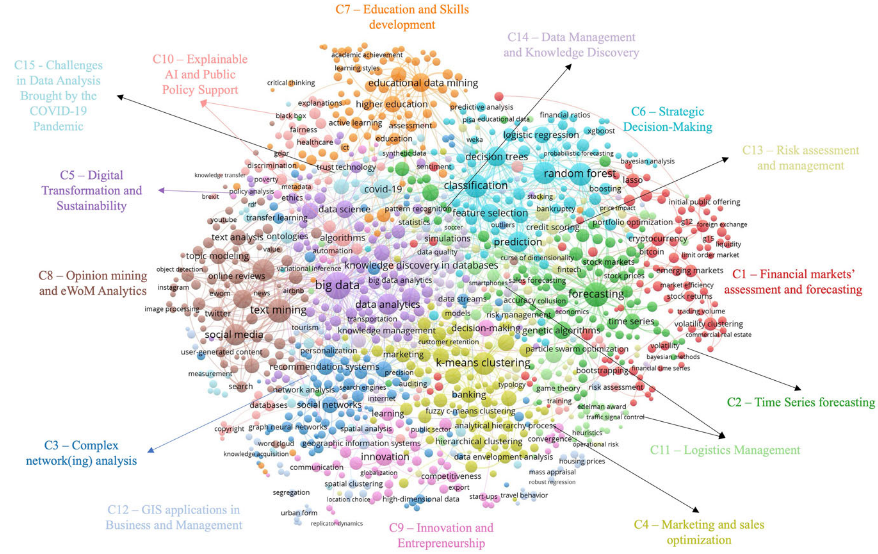


Fig. 3 Co-word network of ML and AI applications in business and management, 2000–mid August, 2022. Unique correlations are hidden to improve legibility. *Note*: The colour of each node denotes the cluster they belong to, and node sizes are proportional to the degree centrality of keywords, that is, the number of correlations each keyword has, both according to the VOS clustering algorithm (van Eck and Waltman, 2007). Source: Author's own elaboration.

analytics; (9) innovation and entrepreneurship; (10) explainable AI and public policy support; (11) logistics management; (12) geographic information systems (GIS) applications in business and management; (13) risk assessment and management; (14) data management and knowledge discovery; and (15) challenges in data analysis due to the COVID-19 pandemic.

C1: Financial markets' assessment and forecasting. Figure 4a explores Cluster 1 (C1), focusing on the application of ML techniques in evaluating and predicting financial assets and market performance. Prominent ML algorithms in this cluster include Linear discriminant analysis (LDA), latent class analyses, and least absolute shrinkage and selection operator (LASSO), which have increasingly supplanted traditional econometric methods such as autoregressive integrated moving average (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH), and Gaussian mixture models (GMM) (Ghiassi et al. 2005; Pai and Lin, 2005; Ahmed et al. 2010; Mullainathan and Spiess, 2017).

Recent developments, especially in cryptocurrency markets like Ethereum and Bitcoin, have seen ML applied to forecast short-term dynamics of these volatile assets, leveraging both historical and real-time data (Urquhart, 2017; Zhang et al. 2018). A notable application is in predicting stock behaviours for bootstrapped enterprises, aiding in portfolio optimization and investment strategies (Manela and Moreira, 2017; Gu et al. 2020).

C2: Time-series forecasting. Figure 4b provides an overview of Cluster 2 (C2), which centres on ML techniques for time-series forecasting. This cluster is pivotal in predicting short-term fluctuations in business metrics such as financial KPIs, including sales and stock prices, and is influenced by macroeconomic variables like unemployment, inflation, and exchange rates (Ghiassi et al. 2005; Ahmed et al. 2010; Ferreira et al. 2016; Manela and Moreira, 2017; Mullainathan and Spiess, 2017; Vidya and Prabheesh, 2020).

Key ML frameworks in this area include long short-term memory (LSTM), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs), which have gained traction for their ability to handle complex temporal data (Law et al. 2019; Fotiadis et al. 2021). GANs, in particular, are noted for their effectiveness in anomaly detection within time series data (Geiger et al. 2020; Zhang et al. 2021). Additionally, traditional algorithms such as particle swarm optimization (PSO), support-vector regression (SVR), ant colony optimization (ACO), and genetic algorithms (GAs) continue to be vital for addressing a wide range of time-series forecasting (Ahmed et al. 2010).

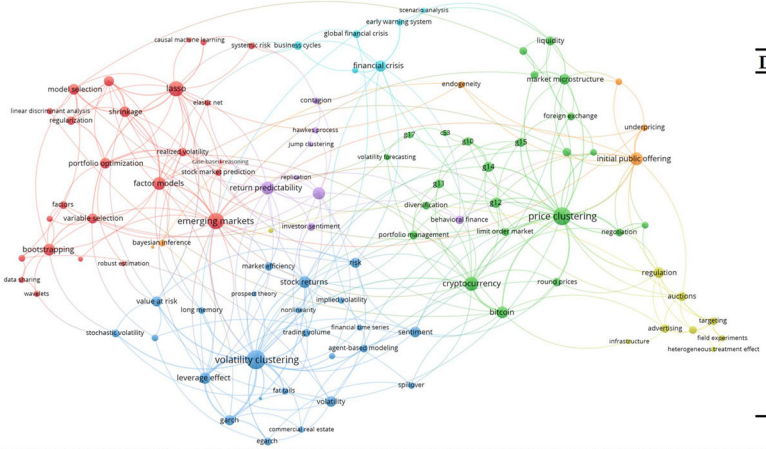
C3: Complex network(ing) analysis. Figure 4c illustrates Cluster 3 (C3), which is primarily focused on complex network analysis. This cluster involves ML techniques rooted in graph theory, including, graph neural networks (GNNs), convolutional neural networks (CNNs) applied to graphs, as well as clustering methods like density-based, grid-based, and affinity propagation clustering.

These methods are applied across diverse areas such as social network analysis (both real-world and online), collaborative filtering, and recommendation systems tailored for e-businesses and e-commerce. In particular, these algorithms support applications like link prediction, user modelling, and adaptive learning, making them crucial for enhancing community-driven recommendations in e-commerce (Wei et al. 2012). Additionally, this cluster includes research into fraud detection using ML techniques and document clustering (Dou et al. 2020).

C4: Marketing and sales optimization. Cluster 4 (C4), depicted in Fig. 5a, focuses on the application of ML in marketing and sales optimization. This cluster centres on the discovery and analysis of market segments and consumer profiles, leveraging clustering techniques and algorithms such as *K*-means, fuzzy *C*-means, and hierarchical clustering methods (Larson et al. 2005). These ML algorithms enhance customer relationship management (CRM) processes by enabling more precise segmentation and profiling of customers.

Within the domain of Segmentation, ML is employed to improve business-to-customer (B2C) and business-to-business (B2B) relationships (McCarty and Hastak, 2007; Paschen et al. 2020a). Techniques like *K*-means clustering and Gaussian mixture models facilitate the customization of high-value-added products and services tailored to specific customer needs, thereby

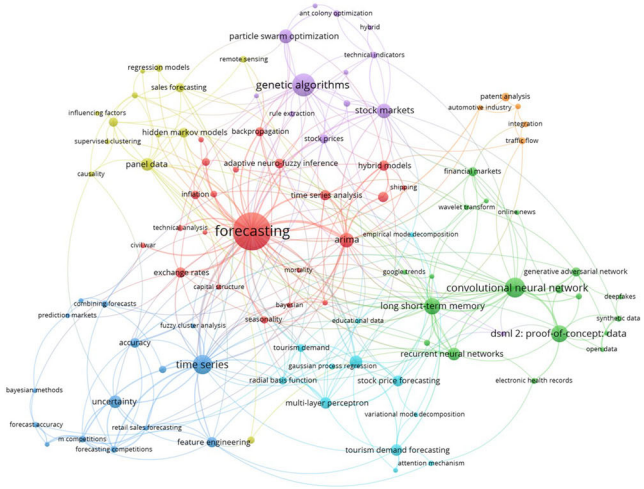
a)



Keywords with the most correlations

Degree	Keyword
44	cryptocurrency
44	lasso
41	volatility clustering
36	emerging markets
34	portfolio optimization
34	price clustering
31	model selection
31	variable selection
29	bitcoin
28	bootstrapping
28	regularization
24	stock returns
23	factor models
23	return predictability
22	asset pricing

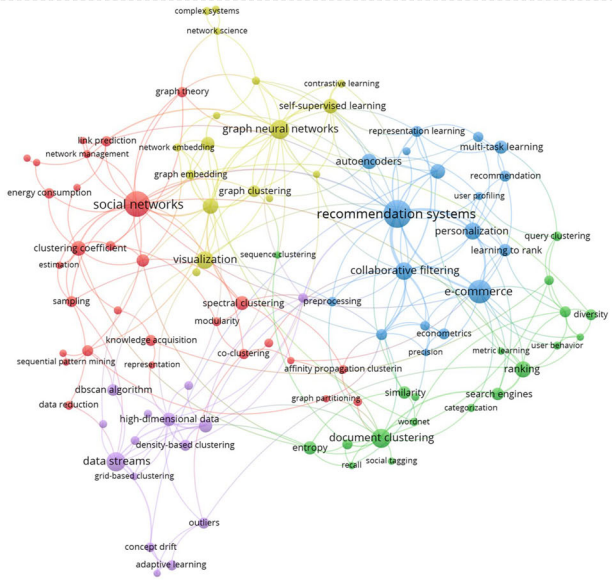
b)



Keywords with the most correlations

Degree	Keyword
189	forecasting
86	genetic algorithms
68	convolutional neural network
65	time series
48	dsml 2: proof-of-concept: data science output has been formulated, implemented, and tested for one domain/problem
46	long short-term memory
41	stock markets
38	multi-layer perceptron
38	uncertainty
34	panel data
33	arima
30	accuracy
29	particle swarm optimization
29	recurrent neural networks
28	feature engineering

c)

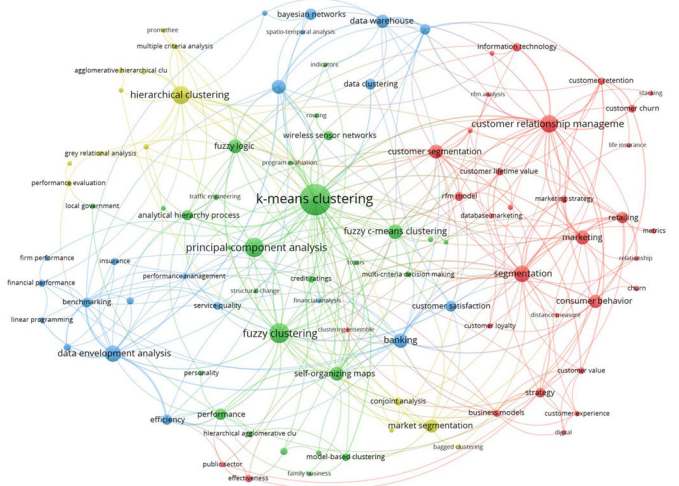


Keywords with the most correlations

Degree	Keyword
73	social networks
72	recommendation systems
64	e-commerce
36	collaborative filtering
36	visualization
34	personalization
32	document clustering
29	data streams
29	ranking
27	transfer learning
26	graph neural networks
24	autoencoders
20	singular value decomposition
19	high-dimensional data
18	community detection

Fig. 4 Zoomed-In View of the Subclustering in Clusters 1, 2 and 3. Detail of **a** Cluster 1: Financial markets' assessment and forecasting. **b** Cluster 2: Time-series forecasting. **c** Cluster 3: Complex network(ing) analysis. *Note*: The colour of each node denotes the cluster they belong to, and node sizes are proportional to the degree centrality of keywords, that is, the number of correlations each keyword has, both according to the VOS clustering algorithm (van Eck and Waltman, 2007). Source: Author's own elaboration.

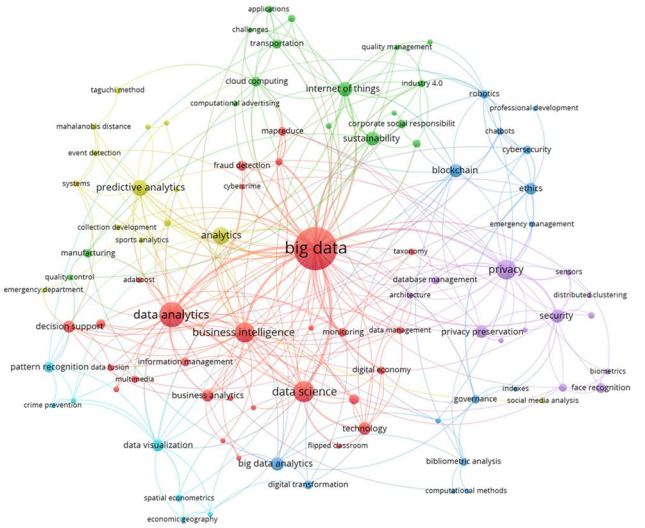
a)



Keywords with the most correlations

Degree	Keyword
160	k-means clustering
68	principal component analysis
64	segmentation
63	banking
62	customer relationship management
61	fuzzy clustering
56	hierarchical clustering
56	marketing
50	decision support systems
45	self-organizing maps
43	fuzzy logic
42	consumer behavior
41	retailing
40	performance
38	data warehouse

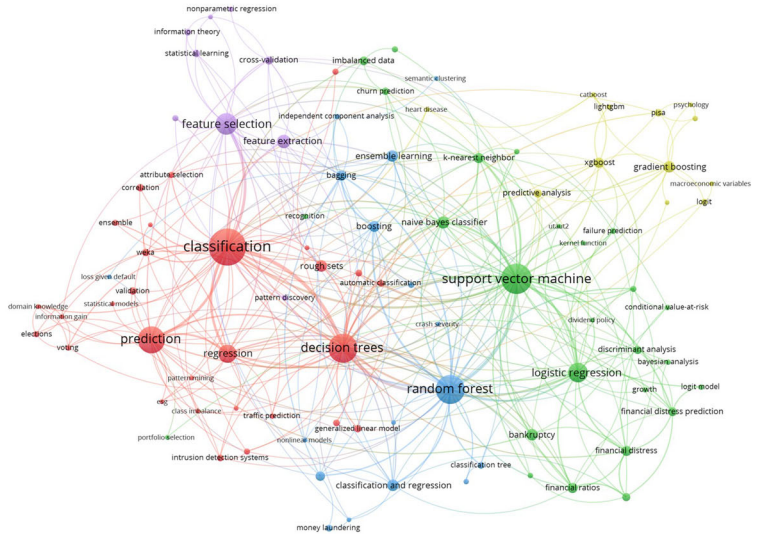
b)



Keywords with the most correlations

Degree	Keyword
167	classification
152	random forest
144	support vector machine
123	prediction
111	decision trees
89	feature selection
73	regression
70	logistic regression
39	naive bayes classifier
37	gradient boosting
34	classification and regression trees
33	ensemble learning
33	k-nearest neighbor
32	boosting
32	regression trees

c)



Keywords with the most correlations

Degree	Keyword
249	big data
129	data analytics
88	business intelligence
86	data science
71	privacy
66	analytics
45	predictive analytics
38	blockchain
31	technology
30	business analytics
30	ethics
30	internet of things
28	big data analytics
27	security
26	decision support

Fig. 5 Zoomed-In View of the Subclustering in Clusters 4, 5 and 6. Detail of **a** Cluster 4: Marketing and sales optimization. **b** Cluster 5: Digital transformation and sustainability. **c** Cluster 6: Strategic decision-making. *Note*: The colour of each node denotes the cluster they belong to, and node sizes are proportional to the degree centrality of keywords, that is, the number of correlations each keyword has, both according to the VOS clustering algorithm (van Eck and Waltman, 2007). Source: Author's own elaboration.

enhancing customer satisfaction and loyalty (Syam and Sharma, 2018). Additionally, ML algorithms support the provision of supplementary services that elevate the overall customer experience.

Another significant application within this cluster is the use of ML for personalized advertising on social media platforms and Google Ads. By leveraging customer profiling through algorithms such as logistic regression, decision trees, and random forests, businesses can deliver more personalized and effective marketing strategies (Lee et al. 2018).

The banking industry prominently utilizes clustering algorithms for customer segmentation and service targeting, distinguishing its ML applications from other sectors that focus on areas like risk assessment, asset management, and stock market analysis (e.g., clusters C1 and C13).

Within C4, ML clustering techniques also complement analytic hierarchical process (AHP), multi-criteria decision-making (MCDM), and decision support systems (DSS) in strategic management. ML algorithms enhance the segmentation, targeting, and positioning (STP) process in marketing by providing data-driven insights that inform more effective segmentation and targeting strategies (Huang and Rust, 2021).

C5: Digital transformation and sustainability. Cluster 5 (C5), illustrated in Fig. 5b, centres on Digital Transformation and Sustainability within the context of Big Data analytics and Business Intelligence (BI). This cluster emphasizes the pivotal role of data-driven decision-making processes in driving enterprise digital transformation (Zaki 2019; Paschen et al. 2020a).

From a management perspective, digital transformation encompasses various Information and Communication Technologies (ICTs) and strategic concepts, including privacy preservation, data analytics, information security, blockchain, and the Internet of Things (IoT). While specific ML algorithms are not the primary focus of this cluster, ML plays a supportive role in several areas. For instance, classification and regression models are used to analyse vast datasets, enabling businesses to derive actionable insights and foster sustainable development (Fuchs et al. 2014; Tabesh et al. 2019; Paschen et al. 2020a). In sustainability efforts, ML contributes through predictive analytics and optimization algorithms that help in creating environmentally conscious business models. Techniques such as regression analysis, time-series forecasting, and unsupervised learning algorithms like principal component analysis (PCA) and clustering are employed to identify patterns and trends that inform sustainable practices (Goralski and Tan 2020; Di Vaio et al. 2020).

Cluster 5 involves high-level strategic discussions on sustainable digital transformation, with ML providing tools and methodologies that enhance data-driven decision-making and support the development of sustainable business models (Zaki 2019; Canhoto and Clear 2020; Di Vaio et al. 2020).

C6: Strategic decision making. The sixth cluster (C6), illustrated in Fig. 5c, delves into strategic decision-making, with a strong emphasis on classification and decision trees. This cluster highlights the use of ML decision tree algorithms and frameworks, including gradient boosting, random forests, and ensemble learning methods, which are pivotal in strategic management contexts (Samoilenko and Osei-Bryson, 2013).

In addition to decision tree-based algorithms, classification algorithms such as support vector machines (SVMs), Naïve Bayes, and *K*-nearest neighbours (KNNs) are extensively employed for strategic decision-making purposes. As defined by some scholars (Ferreira et al. 2016; Bertsimas and Kallus, 2020) these ML methods are instrumental in predictive scenarios, including:

- Forecasting financial churn and distress: Using logistic regression and SVMs to predict customer churn and financial distress.
- Evaluating bankruptcy possibilities: Applying random forests and gradient boosting to assess the likelihood of bankruptcy.
- Identifying failures: Utilizing KNNs and Naïve Bayes classifiers to detect potential failures within business processes or strategies.

These ML algorithms provide robust frameworks for making informed strategic decisions by analysing complex datasets and uncovering patterns that might not be apparent through traditional analytical methods.

C7: Education and skills development. The seventh cluster (C7), depicted in Fig. 6a, focuses on education and skills development. Enterprises are increasingly adopting ML-driven techniques and frameworks for online learning, particularly in training and onboarding procedures. Reinforcement learning algorithms and predictive analytics models are commonly used to personalize learning experiences, enhance assessment processes, and optimize retention rates. These ML methods facilitate performance measurement and increase engagement among trainees and newly onboarded employees, thereby improving the overall onboarding and skills development process (Bakhshinategh et al. 2018; Hellas et al. 2018; Fernandes et al. 2019).

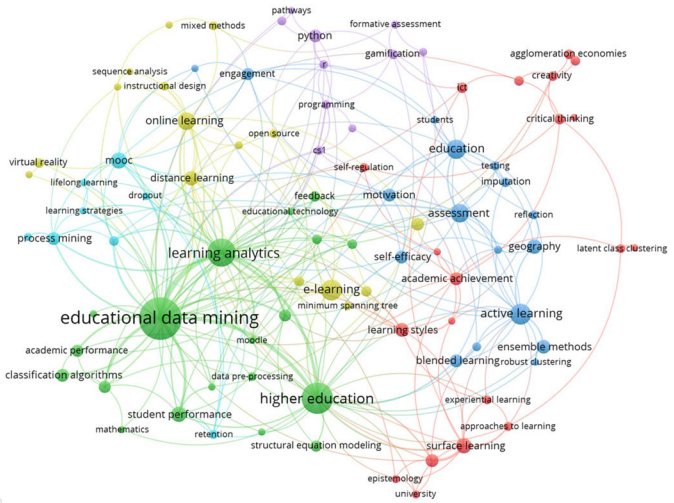
C8: Opinion mining and electronic Word of Mouth (eWOM) analytics. Cluster 8, depicted in Fig. 6b, revolves around opinion mining and electronic Word of Mouth (eWOM) analytics. The maturation of natural language processing (NLP) algorithms has significantly advanced the automated analysis of user interactions on websites, e-commerce platforms, and news reports (Fuchs et al. 2014; Manela and Moreira, 2017; Guo et al. 2017).

ML frameworks for opinion mining frequently utilize algorithms like latent Dirichlet allocation (LDA) for topic modelling and bidirectional encoder representations from transformers (BERT) or Word2Vec for word embeddings. These techniques enable the classification of opinions into pre-defined categories, allowing for the analysis of large volumes of social media content with minimal human intervention (Xiang et al. 2017; Guo et al. 2017; Hartmann et al. 2019). Additionally, content analysis and semantic analysis approaches are enhanced by computer cision techniques, such as image clustering and object detection, to automatically review multimedia content (Ma et al. 2018; Zhang et al. 2019).

C9: Innovation and entrepreneurship. Figure 6c highlights Cluster 9, which centres on Innovation and Entrepreneurial opportunities driven by ML. Although this cluster primarily engages in high-level strategic discourse, without extensive practical applications of ML, there are key areas where ML plays a role. In Financial Technology (FINTECH), enterprises leverage predictive analytics, machine learning-based forecasting models, and decision support systems to navigate rapidly changing market conditions and to innovate in business strategies (Paschen et al. 2020b; Gu et al. 2020).

The use of ML in innovation is also discussed in the context of user experience (UX) design tools (Dove et al. 2017), where algorithms such as reinforcement learning (Rojas-Córdova et al. 2020) and recommendation systems (Graham and Bonner 2022) could potentially enhance user engagement and product

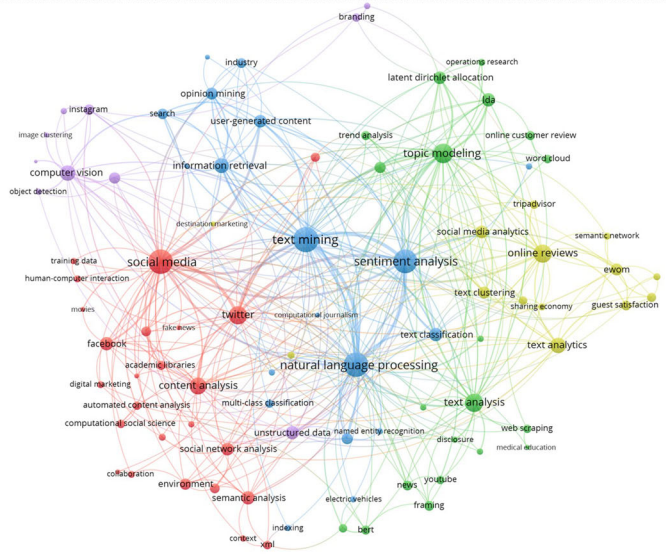
a)



Keywords with the most correlations

Degree	Keyword
95	educational data mining
70	higher education
59	learning analytics
46	e-learning
33	active learning
32	education
31	online learning
29	assessment
28	mooc
23	distance learning
23	python
22	student performance
17	ensemble methods
17	performance prediction
17	r

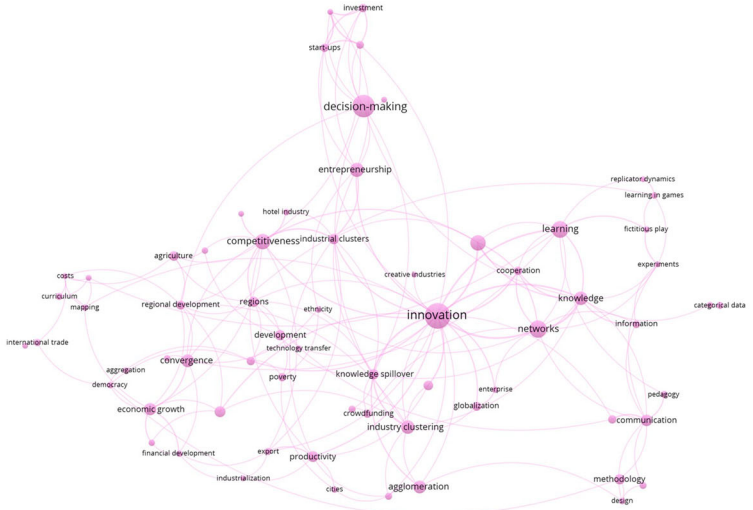
b)



Keywords with the most correlations

Degree	Keyword
179	text mining
145	natural language processing
135	social media
97	sentiment analysis
67	topic modeling
62	text analysis
54	online reviews
52	information retrieval
51	twitter
45	social network analysis
42	text classification
36	computer vision
34	content analysis
29	information extraction
28	search

c)



Keywords with the most correlations

Degree	Keyword
87	innovation
69	decision-making
45	learning
41	networks
37	small and medium-sized enterprises
30	competitiveness
28	knowledge
20	development
20	entrepreneurship
19	economic growth
18	agglomeration
18	communication
17	productivity
16	convergence
15	industry clustering

Fig. 6 Zoomed-In View of the Subclustering in Clusters 7, 8 and 9. Detail of **a** Cluster 7: Education and skills development. **b** Cluster 8: Opinion mining and eWOM analytics. **c** Cluster 9: Innovation and entrepreneurship. Note: The colour of each node denotes the cluster they belong to, and node sizes are proportional to the degree centrality of keywords, that is, the number of correlations each keyword has, both according to the VOS clustering algorithm (van Eck and Waltman, 2007). Source: Author's own elaboration.

development processes. While the exact applications of ML in entrepreneurship remain somewhat ambiguous, ongoing research suggests that ML could play a pivotal role in understanding and forecasting innovation trends and entrepreneurial success.

C10: Explainable AI and public policy support. The tenth cluster (C10), depicted in Fig. 7a, focuses on Explainable AI (XAI) and its critical role in supporting public policies through ML. Researchers in this cluster have developed frameworks to design ML algorithms that are transparent, fair, and high-performing. Key algorithms include SHapley Additive exPlanations (SHAP) and local interpretable model-agnostic explanations (LIME), which help in making ML models more interpretable and understandable by policymakers and the public (Adadi and Berrada, 2018; Preece, 2018; Miller, 2019; Barredo Arrieta et al. 2020). Mitigating biases in ML algorithms is central to XAI, as biases can contribute to public scepticism and distrust in data-driven decision-making. This concern extends to various sectors, including businesses, where biases impact employees and stakeholders (Kaplan and Haenlein, 2020; Paschen et al. 2020b). The ethical implications of ML algorithms, particularly regarding the potential reinforcement of biases against marginalized groups, have led to discussions on the responsible use of ML. Responses to these ethical challenges include regulatory frameworks like the European Union's General Data Protection Regulation (GDPR) and the ongoing development of XAI methods (Preece, 2018; Kaplan and Haenlein, 2020; Canhoto and Clear, 2020).

C11: Logistics management. The eleventh cluster (C11), illustrated in Fig. 7b, examines ML applications in logistics management, emphasising the optimisation of operations and cost reduction. Key ML frameworks in this cluster include deep reinforcement learning and multi-agent systems, which are applied to solve complex logistics problems such as vehicle routing problems (VRP) and travelling salesman problems (TSP) (Chen and Wu, 2005; Luo et al. 2009; Ferreira et al. 2016).

Traditional optimization methods like simulated annealing and Tabu search are increasingly being complemented or replaced by advanced ML techniques, including spatial–temporal clustering and debiased ML. These methods are critical in supply chain management (SCM) tasks, such as identifying optimal delivery routes, forecasting product demand, agile supplier selection, dynamic pricing, and scheduling. By incorporating ML, businesses can enhance their logistics efficiency and reduce operational costs (Huang et al. 2011).

C12: Geographic information systems (GIS) applications in business and management. Cluster 12 (C12), depicted in Fig. 7c, is dedicated to the integration of geographic information systems (GIS) with ML techniques to enhance business and management applications. This cluster includes the use of spatial clustering and spatial data mining algorithms to develop and coordinate multimodal supply routes, assess housing prices for optimal location selection, and re-evaluate business investment values (Grekousis et al. 2013; Chen and Tsai, 2016). Beyond traditional GIS applications, this cluster explores the impact of criminal and gang activity, traffic patterns, and real estate dynamics on business operations. By integrating GIS with ML, such as *K*-means clustering and density-based spatial clustering of applications with noise (DBSCAN), businesses gain valuable insights that inform decision-making and improve operational efficiency. These ML-driven GIS applications are crucial for strategic planning, including workforce performance management, recruitment, onboarding processes, and asset valuation (Morency et al. 2007; Chang et al. 2010; Grekousis et al. 2013).

C13: Risk assessment and management. Cluster 13 (C13), depicted in Fig. 8a, focuses on risk assessment and management, particularly within the financial sector. ML plays a critical role in this domain, especially in assessing customer credit risk and managing debt risks. Traditional methods, such as Monte Carlo simulations, are increasingly being complemented by advanced ML algorithms like random forests, support vector machines (SVMs), and neural networks (Khandani et al. 2010; Bekhet and Eletter, 2014; Koutanaei et al. 2015).

This cluster extends beyond financial risks, encompassing business risks in project management, fraud detection, auditing, and investments, particularly in emerging economies. ML-driven techniques such as anomaly detection and predictive analytics are widely used in these areas to identify and mitigate potential risks (Thiprungsri and Vasarhelyi, 2011; Jin and Zhang, 2011; Gray and Debreceny, 2014). The FINTECH industry's influence is notable, relying heavily on ML and ICTs to offer innovative financial services, although specific ML applications within FINTECH are often intertwined with those in clusters like C1 and C2 (Kalinic et al. 2019; Gu et al. 2020).

C14: Data management and knowledge discovery. The cluster depicted in Fig. 8b (C14) focuses on data management and knowledge discovery, a critical area for ML applications in business. This cluster highlights the processes involved in acquiring, preprocessing, and analysing data to maximize business value. The cross-industry standard process for data mining (CRISP-DM) remains the dominant framework in this area, guiding the knowledge discovery in databases (KDD) process (Fernandes et al. 2019).

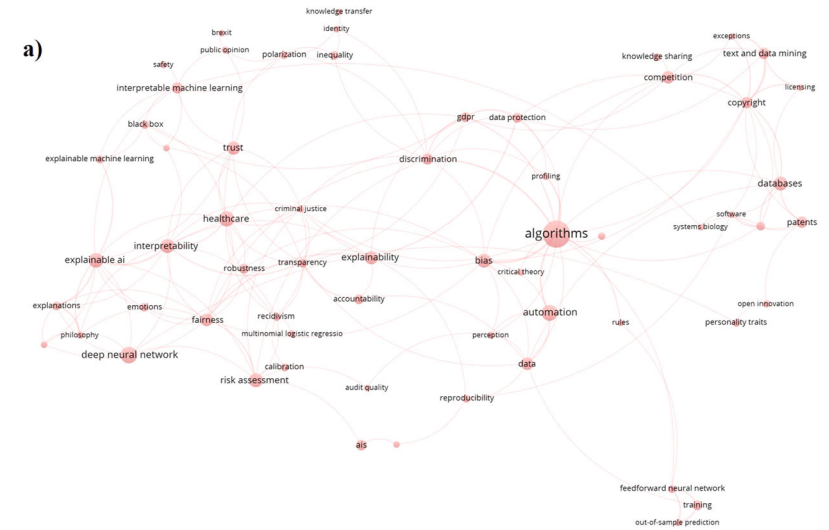
ML techniques central to this cluster include constrained clustering, functional clustering, restricted Boltzmann machines, latent semantic indexing (LSI), and a priori algorithms. These algorithms are applied to continuous data streams from sources like web content, surveys, and market baskets, enabling businesses to perform tasks such as data cleaning, outlier detection, and data imputation (Chen and Wu, 2005). Additionally, non-ML techniques like linear regressions, Markov chains, and conceptual modelling complement these ML methods to ensure data quality and privacy in business operations.

C15: Challenges in data analysis brought by the COVID-19 pandemic. The last cluster, C15, centres on the challenges in data analysis brought by the COVID-19 pandemic, as depicted in Fig. 8c. ML algorithms have been pivotal in predicting the pandemic's impact on businesses, with models like time-series forecasting and Bayesian networks being employed to assess the effects of lockdowns and restrictions (Vidya and Prabheesh, 2020; Baek et al. 2020; Fotiadis et al. 2021).

The pandemic has accelerated digital transformation across various industries, pushing businesses towards extensive digitization and the adoption of remote work policies. These changes have created new opportunities for ML applications in digital transformation projects, particularly those detailed in Cluster C5. Semantic networks and meta-analysis techniques, along with bibliometric analysis, have become essential tools for businesses navigating this new landscape, facilitating their transition to more digital operations (Sedera et al. 2022; Miklosik and Krah, 2023).

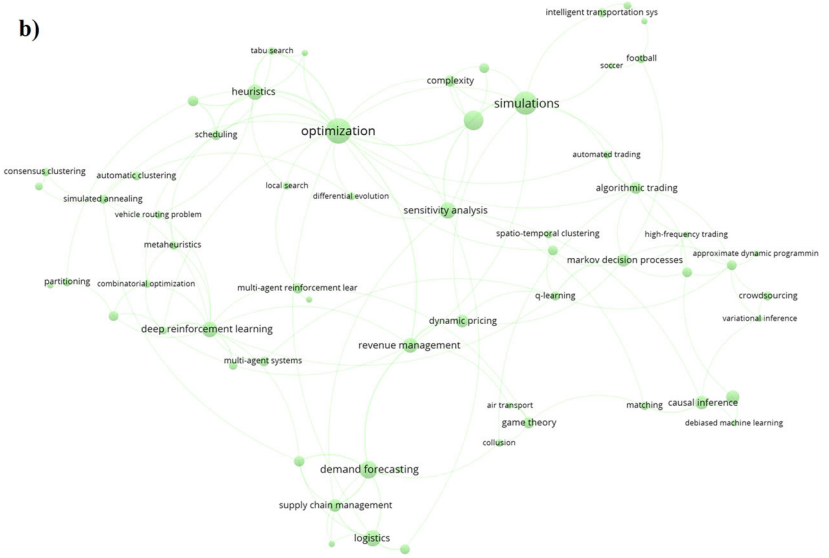
Discussion

Analysis from the strategic diagram. Machine learning (ML) is increasingly central to business administration, with applications



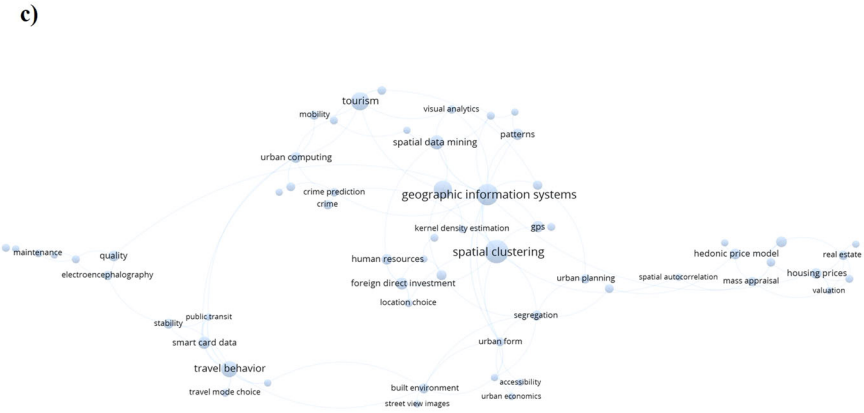
Keywords with the most correlations

Degree	Keyword
81	algorithms
36	bias
31	deep neural network
30	databases
29	explainable ai
28	interpretable machine learning
27	data
27	trust
26	healthcare
25	automation
24	fairness
24	interpretability
20	competition
20	risk assessment
18	transparency



Keywords with the most correlations

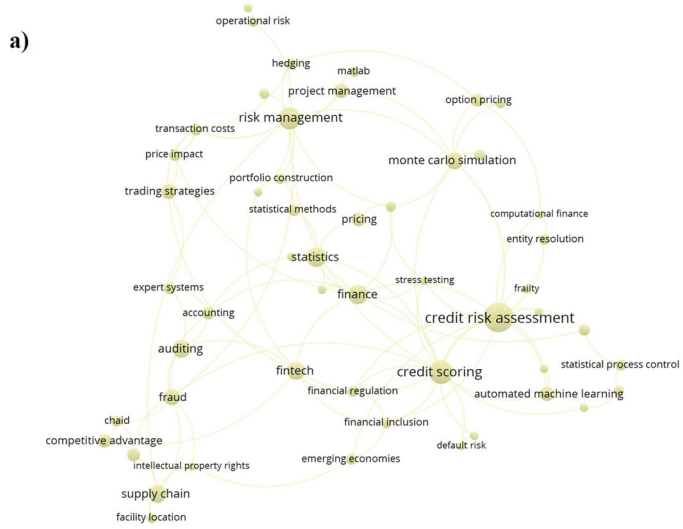
Degree	Keyword
69	optimization
50	simulations
49	modeling
36	sensitivity analysis
29	demand forecasting
27	revenue management
26	logistics
23	game theory
22	supply chain management
19	algorithmic trading
19	causal inference
17	complexity
17	heterogeneity
16	deep reinforcement learning
15	gaussian processes



Keywords with the most correlations

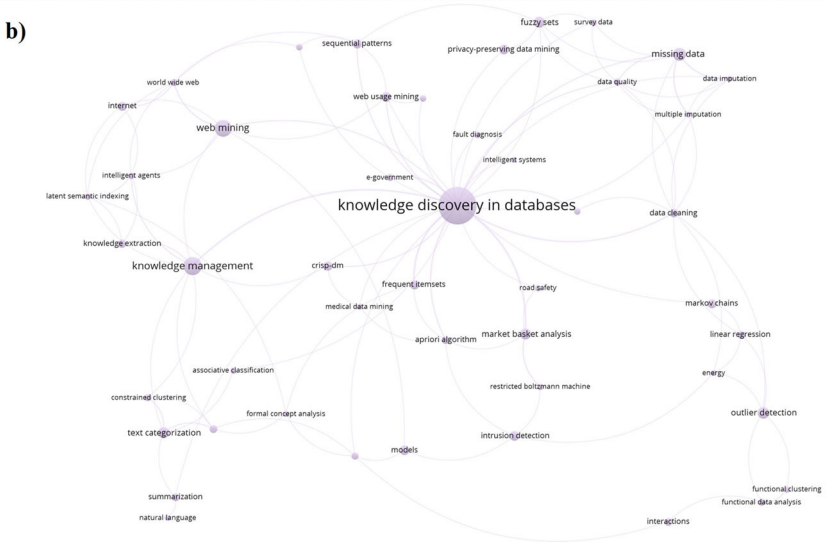
Degree	Keyword
28	geographic information systems
26	tourism
21	spatial analysis
19	spatial clustering
19	stability
15	travel behavior
14	foreign direct investment
14	spatial data mining
12	housing prices
12	human resources
12	mobility
11	location
11	quality
11	real estate
11	segregation

Fig. 7 Zoomed-In View of the Subclustering in Clusters 10, 11 and 12. Detail of **a** Cluster 10: Explainable AI and public policy support. **b** Cluster 11: Logistics management. **c** Detail of cluster 12: GIS applications in business and management. *Note*: The colour of each node denotes the cluster they belong to, and node sizes are proportional to the degree centrality of keywords, that is, the number of correlations each keyword has, both according to the VOS clustering algorithm (van Eck and Waltman, 2007). Source: Author's own elaboration.



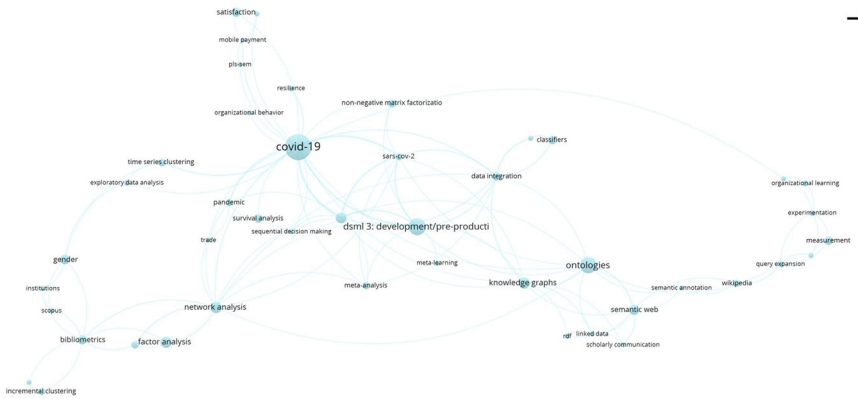
Keywords with the most correlations

Degree	Keyword
56	credit risk assessment
43	credit scoring
40	finance
32	risk management
28	statistics
28	supply chain
22	fintech
21	monte carlo simulation
20	auditing
19	trading strategies
16	fraud
16	risk analysis
14	accounting
13	automated machine learning
13	transaction costs



Keywords with the most correlations

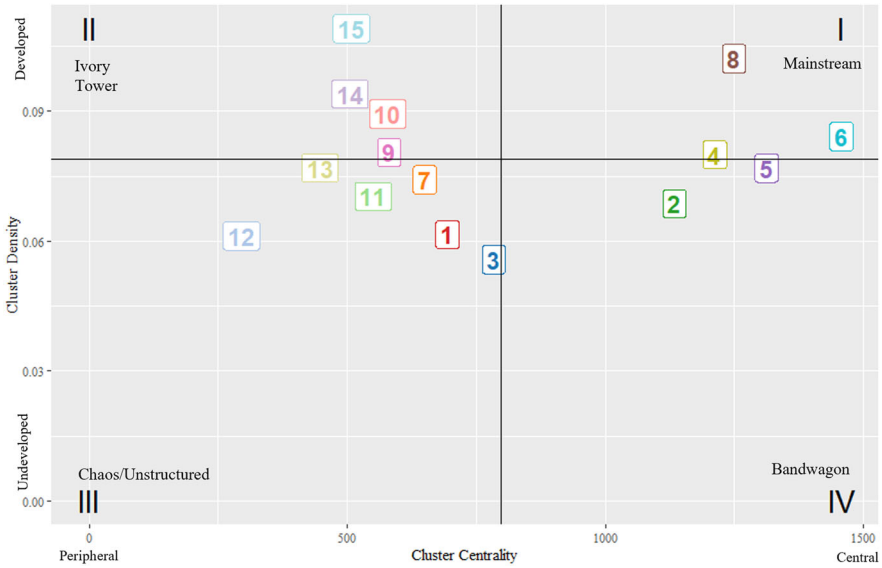
Degree	Keyword
115	knowledge discovery in databases
47	knowledge management
44	web mining
34	missing data
21	data cleaning
19	internet
18	fuzzy sets
18	knowledge extraction
18	models
16	data quality
15	outlier detection
14	text categorization
14	web usage mining
13	conceptual modeling
13	digital libraries



Keywords with the most correlations

Degree	Keyword
111	covid-19
61	dsml 3: development/pre-production: data science output has been rolled out/validated across multiple domains/problems
44	ontologies
36	network analysis
25	gender
24	knowledge graphs
23	coronavirus
21	factor analysis
20	semantic web
18	classifiers
17	bibliometrics
15	time series clustering
14	data integration
14	meta-analysis
14	sars-cov-2

Fig. 8 Zoomed-In View of the Subclustering in Clusters 13, 14 and 15. Detail of **a** Cluster 13: Risk assessment and management. **b** Cluster 14: Data management and knowledge discovery. **c** Detail of Cluster 15: Challenges in data analysis brought by COVID-19. *Note*: The colour of each node denotes the cluster they belong to, and node sizes are proportional to the degree centrality of keywords, that is, the number of correlations each keyword has, both according to the VOS clustering algorithm (van Eck and Waltman, 2007). Source: Author's own elaboration.



Cluster	Name or Description
1	Financial Markets' assessment and forecasting
2	Forecasting macroeconomic and performance indicators
3	Complex Network Analysis
4	Marketing and Sales Optimization
5	Digital Transformation and Sustainability
6	Strategic Decision-Making
7	Education and Skills Development
8	Opinion Mining and eWOM Analytics
9	Innovation and Entrepreneurship
10	Explainable AI and Public Policy Support
11	Logistics Management
12	GIS in Business and Management
13	Risk assessment and management
14	Data Management and Knowledge Discovery
15	Challenges in Data Analysis brought by COVID-19

Fig. 9 Strategic diagram of ML and AI applications in business and management. Source: Author's own elaboration.

Table 1 Cluster metrics for ML applications in business and management.

Cluster	Nodes	Edges	Centrality	Density
C1—Financial markets' assessment and forecasting	102	317	694	0.06154
C2—Time series forecasting	99	333	1134	0.06865
C3—Complex network(ing) analysis	98	265	784	0.05575
C4—Marketing and sales optimization	95	357	1212	0.07996
C5—Digital transformation and sustainability	94	336	1312	0.07687
C6—Strategic decision-making	93	360	1457	0.08415
C7—Education and skills development	90	297	650	0.07416
C8—Opinion mining and eWoM analytics	88	391	1249	0.10214
C9—Innovation and entrepreneurship	67	178	581	0.08051
C10—Explainable AI and public policy support	62	169	578	0.08937
C11—Logistics management	62	133	550	0.07033
C12—Geographic information systems applications in business and management	61	112	295	0.06120
C13—Risk management and assessment	52	102	447	0.07692
C14—Data management and knowledge discovery	48	106	505	0.09397
C15—Challenges in data analysis brought by the COVID-19 pandemic	45	108	508	0.10909

Source: Author's own elaboration.

that span a wide array of domains. A co-word network analysis revealed fifteen distinct clusters, each representing a unique area where ML techniques are applied. These clusters, visualized through a strategic diagram (Fig. 9), highlight the diverse and interconnected landscape of ML in business and management. The strategic diagram further categorizes these clusters based on centrality and density metrics, offering insights into their development and potential impact.

A pivotal aspect of this analysis is the strategic diagram delineating ML applications in business and management, which visualizes centrality and density metrics within keyword clusters on a two-dimensional graph as described in Table 1. Centrality measures the sum of correlation frequencies or 'link strengths' between cluster keywords and those elsewhere in the co-word map (Callon et al. 1991). Meanwhile, cluster density gauges the ratio of cluster keywords to their correlations, signifying the conceptual development within a theme (Callon et al. 1991).

Advanced and sought-after applications. The strategic diagram identifies marketing and sales optimization (C4), opinion mining and eWOM analytics (C8), and strategic decision-making (C6) as the most advanced and sought-after ML applications, located in Quadrant I (mainstream). These clusters demonstrate mature use cases with ongoing research and practical application. For instance, ML's role in marketing and sales optimization is evident in the refinement of customer segmentation and targeted advertising strategies, enhancing customer relationship management (CRM). Similarly, in strategic decision-making, ML techniques like decision trees and ensemble learning are crucial for predictive analytics, such as forecasting financial distress and bankruptcy.

Emerging and promising themes. Emerging themes like time-series forecasting (C2) and digital transformation and sustainability (C5), found in Quadrant IV (Bandwagon), indicate their growing relevance and potential for broader adoption. ML's significance in time-series forecasting is underscored by its application in predicting short-term business metrics, such as sales and stock prices, which are crucial for strategic planning. The diversity of ML frameworks, from LSTM networks to particle swarm optimization, illustrates ML's adaptability in addressing various forecasting challenges across industries. Meanwhile, digital transformation and sustainability are

becoming increasingly vital as businesses shift towards data-driven decision-making, integrating advanced ICTs for sustainable development.

Clusters situated in Quadrant III (chaos/unstructured), such as risk management (C13), logistics management (C11), and complex network analysis (C3), show promise for future research. Risk management, for example, benefits from ML's capabilities in assessing financial risks and optimizing investment strategies. In logistics, ML enhances supply chain efficiency through advanced optimization and simulation methods. The application of ML in complex network analysis, particularly in social networks and recommendation systems, demonstrates how businesses can leverage community behaviours to refine marketing strategies and customer interactions.

Challenges and ethical considerations. Explainable AI and public policy support (C10), positioned in Quadrant II (Ivory Tower), emphasize the importance of transparency, fairness, and trust in ML applications. The growing focus on explainable AI (XAI) reflects the need to ensure that ML algorithms are not only effective but also ethically sound, particularly in public policy and business decision-making. Ethical considerations, such as mitigating biases in algorithmic outcomes, are critical to maintaining public trust and ensuring that ML's deployment in sensitive areas, like financial markets or social services, is both fair and just.

Evolving and declining relevance. Certain themes, like innovation and entrepreneurship (C9) and data management and knowledge discovery (C14), although historically significant, are seeing a decline in relevance as their use cases mature. Positioned in Quadrant II, these themes have garnered substantial academic attention but are becoming less central to current applied research. For instance, the challenges in data analysis due to COVID-19 (C15) are becoming less relevant as businesses adapt to the post-pandemic landscape. However, there remains potential to integrate these maturing themes with more dynamic ones, such as linking knowledge discovery to strategic decision-making, to sustain their relevance in the rapidly evolving field of business ML applications.

Discussion on machine learning applications in various domains of business administration. Machine learning (ML) has become integral to modern business administration, providing solutions that enhance efficiency, decision-making, and customer satisfaction across various sectors. Thus, as illustrated in Table 2, we delve into the various algorithms to uncover their practical applications in the fields of business and administration.

Finance (Clusters C1, C2, C13). In the financial sector, ML algorithms are employed extensively for risk management and market forecasting. For example, linear discriminant analysis and latent class analysis are used to predict stock market returns and identify high-risk investments, particularly in volatile markets like cryptocurrencies. Companies use these techniques to optimize their investment portfolios, ensuring better returns by balancing risk and reward. Additionally, genetic algorithms and recurrent neural networks (RNNs) are applied in asset price forecasting, allowing financial institutions to make informed decisions on stock purchases and sales. In risk assessment, automated machine learning helps banks evaluate creditworthiness and detect potential fraud, significantly reducing financial losses and improving credit scoring processes.

Customer relationship management (Clusters C3, C4, C8). ML plays a crucial role in enhancing customer relationship management (CRM). Graph neural networks are leveraged in e-commerce platforms to create personalized shopping experiences. For instance, Amazon uses recommendation systems that analyse user behaviour and purchase history to suggest products, increasing customer satisfaction and sales. In marketing, *K*-means clustering is utilized to segment customers based on purchasing habits, allowing businesses to tailor marketing campaigns effectively. Natural language processing (NLP) and latent Dirichlet allocation (LDA) are employed in sentiment analysis to monitor social media platforms and customer reviews, helping companies like Starbucks to adapt their products and services based on real-time customer feedback.

Decision-making support (Clusters C6, C11). Strategic decision-making is another area where ML has significant applications. Random forests and support vector machines (SVMs) assist companies in making data-driven decisions. For example, in logistics, businesses like FedEx use spatial–temporal clustering to optimize delivery routes, reducing transportation costs and improving delivery times. In revenue management, airlines like Delta Airlines use multi-agent reinforcement learning to dynamically adjust ticket prices based on demand, maximizing profits while maintaining customer satisfaction.

Innovation and public policy (Clusters C7, C9, C10). In the field of education and corporate training, latent class clustering is used to personalize learning experiences, improving outcomes in online learning platforms such as Coursera. For public policy support, governments use explainable AI to ensure transparency in automated decision-making, especially in areas like criminal justice where bias mitigation is critical. For instance, public health agencies during the COVID-19 pandemic utilized multinomial logistic regression to predict infection spread and guide policy decisions, such as lockdown measures.

Data management and sustainability (Clusters C14, C5, C12, C15). ML-driven data management is vital for businesses seeking to extract meaningful insights from vast datasets. Latent semantic indexing (LSI) is used by companies like Google to enhance search engine capabilities, ensuring users receive the most relevant results. In the real estate sector, spatial data mining helps companies analyse location data to determine optimal sites for new developments, considering factors like housing prices and traffic patterns. During the COVID-19 pandemic, time-series clustering enabled businesses to track the impact of the virus on supply chains, allowing companies like Walmart to adjust their logistics strategies and maintain product availability.

Conclusion

This study offers an understanding of the widespread integration of machine learning (ML) across diverse domains within business and management. Our analysis uncovers a distinct approach to employing ML, diverging from the usual classifications found in the literature, which typically highlight four typical business domains: Finance, Marketing, Firm Strategy, and Human Resource Management (Zhang et al. 2021). Hence, this work presents a comprehensive taxonomy, delineating fifteen thematic clusters. Detailed applications of ML in various business contexts, systematically presented in Table 2, categorize thematic clusters and the specific ML algorithms, methods, or techniques employed within each context. This comprehensive resource, a synthesis of various studies, provides a valuable reference point for managers confronting challenges aligned with each thematic cluster. It enables them to explore cited algorithms and technologies, offering insights and potential strategies tailored to their specific managerial domains.

Table 2 Most common algorithms and their business application.

Business application	Cluster	Thematic involved	Most used algorithms and technologies
Finance	C1—Assessing and forecasting the performance of financial markets	Price clustering Volatility Return predictability Cryptocurrency and currency markets (FOREX) Portfolio optimization and management Investments in emerging economies	Linear discriminant analysis Latent class analysis Causal machine learning Agent-based modelling
	C2—Forecasting macro- and micro-economic indicators	Stock market forecasting Demand and sales forecasting Asset price forecasting	Genetic algorithms Particle swarm optimization Long short-term memory Recurrent neural networks Generative adversarial networks
	C13—Risk assessment and management	Credit scoring Credit risk assessment Fraud Stress testing Default risk	Automated machine learning Chi-square automatic interaction detection (CHAID)
Customer relationship management	C3—Complex network(ing) Analysis	Social networking E-commerce recommendation systems Personalization Document clustering	Graph neural networks Autoencoders Density-based clustering algorithm (DBSCAN) Spectral clustering Affinity Affinity propagation clustering
	C4—Marketing and sales optimization	Customer satisfaction and loyalty Customer retention and churn Benchmarking Segmentation, targeting and positioning Recency, frequency, monetary value (RFM) models and analysis	*K*-means clustering Principal component analysis Fuzzy [*C*-means] clustering
	C8—Opinion mining and eWOM analysis	Customer reviews User-generated content Social media Sentiment analysis Information retrieval	Natural language processing Topic modelling Latent Dirichlet allocation Computer vision
Decision-making support	C6—Strategic decision-making	Classification Prediction Regression analysis Decision theory	Random forests Support vector machines Gradient boosting Naïve Bayes classifiers *K*-nearest neighbours
	C-11—Logistics management	Demand forecasting Revenue management Intelligent transportation systems Game theory Markov decision processes	Spatio-temporal clustering Multi-agent reinforcement learning Debiased machine learning
Innovation	C7—Education and skills development	Learning analytics Academic performance Online learning and e-Learning Distance learning Critical thinking Gamification	Latent class clustering Robust clustering Minimum spanning trees Educational data mining
	C9—Innovation and entrepreneurship	Decision-making Competitiveness Entrepreneurship Knowledge spillover	
	C14—Data management and knowledge discovery	Knowledge management Market basket analysis Knowledge discovery in databases	Intelligent systems Latent semantic indexing Web mining Cross-industry standard process for data mining (CRISP-DM)
Public Policy Support	C10—Explainable AI and public policy support	Explainability Accountability Profiling Biases	Explainable AI Deep neutral networks Multinomial logistic regression

Table 2 (continued)

Business application	Cluster	Thematic involved	Most used algorithms and technologies
		Trust Fairness	
Other topics and applications	C5—Digital transformation and Sustainability	Business intelligence Predictive analytics Decision support Business analytics Industry 4.0	MapReduce AdaBoost Distributed clustering Distributed data mining
	C12—GIS applications in Business and Management	Location choice Housing prices Mass appraisal Travel behaviour & mode choice	Spatial clustering Spatial data mining Visual analytics
	C15—Challenges in data analysis brought by COVID-19	Ontologies Data integration	Time-series clustering Semantic web Knowledge graphs

Source: Author's own elaboration.

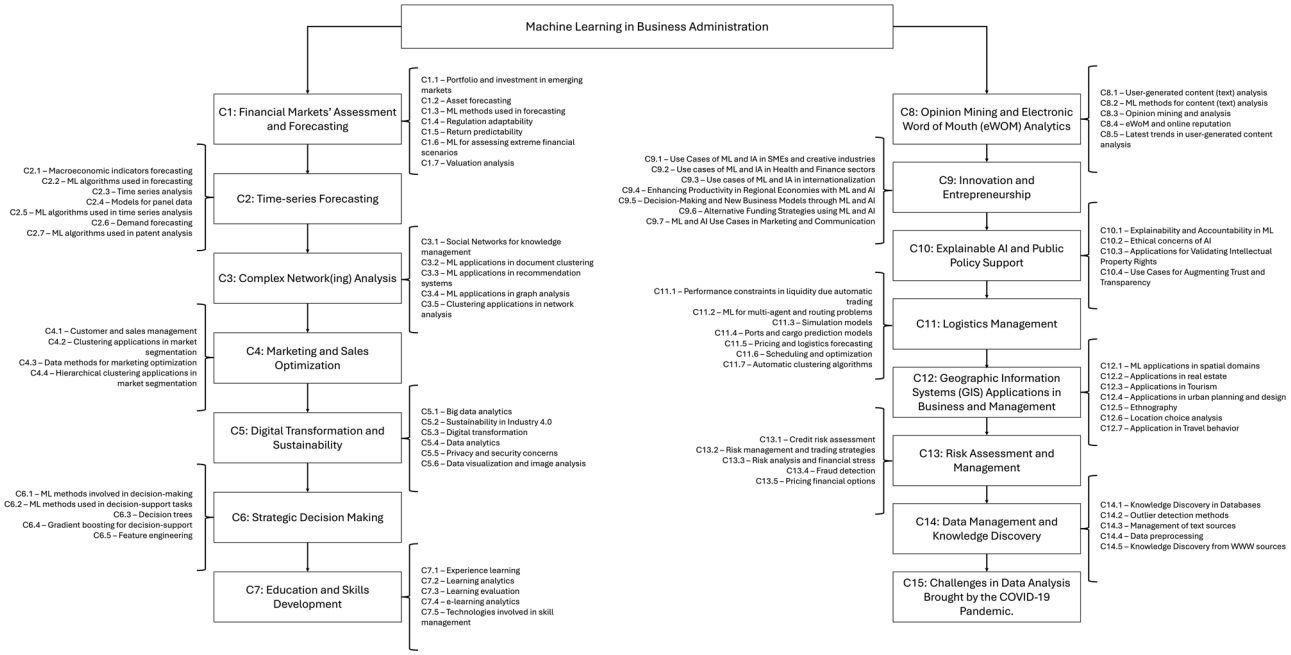


Fig. 10 A taxonomy of ML and AI applications in business and management. Source: Author's own elaboration.

Moreover, based on the analysis and classification of clusters in business and administration, it is evident that machine learning (ML) and artificial intelligence (AI) are increasingly integral to optimizing various domains, from financial markets to education, logistics, and public policy. Each cluster highlights the growing reliance on ML algorithms to enhance decision-making, risk management, and operational efficiency across industries. For instance, financial markets benefit from advanced forecasting techniques, while logistics management is revolutionized through ML-driven optimization and scheduling. Furthermore, the integration of ML in education and skills development underscores the technology's potential in personalized learning and performance assessment.

The taxonomy of clusters illustrated in Fig. 10, displays the diverse applications of ML and AI, demonstrating their transformative impact on traditional business processes. As these technologies continue to evolve, their ability to provide actionable insights, streamline operations, and support strategic decisions will be paramount. The clusters reveal a holistic view of how ML and AI are reshaping business landscapes, offering a roadmap for future research and practical implementations that can drive innovation and sustainable growth across sectors.

Recommendations for Managers. In the realm of machine learning (ML) and intelligent automation (IA) technologies, managerial roles play a pivotal role in conducting thorough evaluations tailored to specific business cases (Tabesh et al. 2019; Miklosik and Krah, 2023). It is crucial to allocate resources while mitigating potential impacts on social, labour, financial, and regulatory facets inherent in the implementation of these technologies (Miller, 2019; Miklosik and Krah, 2023). Hasty adoption of ML or IA, driven by industry trends or competitive pressures, may result in unintended consequences beyond immediate financial implications.

It is recommended avoiding extreme decisions in ML adoption (Tabesh et al. 2019). The implementation of ML technologies should not entail outright replacement of existing structures or

paradigms, nor should it lead to a wholesale rejection of transformative tools. Instead, a nuanced, gradual, and consensus-driven approach is recommended. This involves strategically and ambitiously integrating ML into existing frameworks, recognizing that embracing this technology demands careful calibration.

The caution extends beyond financial considerations, encompassing awareness of the dynamic shifts in popularity and the increased regulatory scrutiny directed at specific technologies such as ChatGPT and large language models (LLMs). Managers are strongly advised to transcend trends, directing their attention to the specific needs of their industry and sector. This strategic focus ensures the enduring sustainability and resilience of their chosen machine learning (ML) strategies, fostering adaptability in the face of evolving technological and regulatory landscapes.

Finally, the judicious incorporation of ML can enhance profitability and competitiveness (Luo et al. 2009; Zaki, 2019; Sedera et al. 2022). Striking a balance between ambition and realism ensures technological advancements bolster productivity without rendering individuals obsolete. A strategic ML deployment should empower individuals within the organization, fostering a sense of support and collaboration, rather than instigating anxiety about being replaced.

Recommendations for developers and academia. Table 2 illustrates the varied applications of ML within these domains, highlighting its significance in contemporary business landscapes. However, despite its widespread integration, persistent challenges hinder the complete realization of ML algorithms' potential.

First, an ongoing area of enquiry revolves around the comparative performance of heuristic models versus specific ML algorithms. Questions remain regarding whether heuristic models might outperform ML algorithms in certain scenarios, particularly given concerns about the potentially slow convergence of ML algorithms within complex systems. This enquiry calls for a deeper analysis of contexts where heuristic approaches could provide more efficient solutions and the optimal scenarios for deploying ML algorithms to their fullest potential. Additionally, further research should focus on evaluating the efficiency and effectiveness of various AI and ML algorithms across different domains of business management, recognizing that AI and ML may not always be the most suitable or effective solutions for some of the challenges faced in this field. Moreover, although in certain domains, ML continues to demonstrate efficacy, year-over-year improvement in most benchmarks is marginal (Vial, 2019; Sedera et al. 2022).

Second, beyond algorithmic performance, the underpinning infrastructure poses critical challenges in realizing the complete implementation of ML algorithms. Infrastructure considerations encompass diverse facets, including computational capacity, processing capabilities, and security protocols (Sullivan et al. 2001; Canhoto and Clear, 2020). Addressing these infrastructure challenges becomes imperative for the seamless integration and effective utilization of ML algorithms within real-world business contexts.

Third, the pragmatic aspects of cost-effectiveness and the challenges associated with the practical integration of these algorithms within existing business frameworks emerge as focal points for further exploration. Understanding the economic viability of implementing ML solutions and navigating the hurdles inherent in their seamless integration into operational systems represent critical aspects demanding closer scrutiny.

Fourth, ML algorithms not only integrate into established businesses across various fields but also play a pivotal role in fostering new enterprises and innovative business models (Atlantico, 2023). However, addressing the digital divide between small and medium-sized enterprises (SMEs) and larger corporations is crucial. Further research should be directed toward evaluating the effectiveness of public policies in supporting SMEs to manage the transitions involved in adopting ML and AI. Additionally, it should explore the potential for SMEs to leapfrog their larger counterparts by strategically adopting these technologies. Addressing these challenges necessitates collaborative efforts among scholars and practitioners across disciplines, especially to support the needs of SMEs. Designing ML algorithms and AI solutions that are responsive to real-world market demands requires considering a broader spectrum of resources, elements, processes, and products. This also involves examining various cooperation levels within business and management processes, particularly those involving multiple agents, to ensure comprehensive and effective implementation.

Certain thematic clusters, like financial markets' assessment and forecasting (C1), complex network analysis (C3), education and skill development (C7), and logistics management (C11), have garnered significant academic interest. Notably, recent startups, especially in developing regions, focus on addressing challenges within these clusters by incorporating ML algorithms into their solutions (Atlantico, 2023). This trend indicates the growing gap between industry and academia in ML algorithm development, as observed in recent reports (Atlantico, 2023; Stanford "AI Index Report", 2023).

Finally, the study acknowledges the crucial role of human interactions, both in the development and utilization of these algorithms. The human element, whether in designing, refining, or utilizing ML algorithms, significantly influences their effectiveness and ethical considerations (Miller, 2019; Miklosik and Krah, 2023). Future research should focus on examining the interplay between human involvement and machine-driven decision-making processes to ensure the responsible and effective application of machine learning (ML) in business and management contexts. Furthermore, the adoption and development of these algorithms necessitate additional skill development among employees and structural changes within businesses, impacting operational models, profit structures, and other fundamental aspects of business frameworks. Closing this gap will be essential in ensuring a more inclusive and effective integration of ML into diverse business landscapes.

Recommendations for policymakers. Based on our analysis, there are significant public policy implications regarding the integration of machine learning (ML) in business and management. Policymakers should prioritize the development of regulatory frameworks that address ethical considerations such as transparency, fairness, and bias mitigation in ML applications. This is particularly crucial in sectors like finance, healthcare, and public services, where algorithmic decisions can significantly impact societal well-being. Regulations should be oriented to promote the use of Explainable AI to ensure that ML algorithms are understandable and accountable, thereby fostering public trust and preventing discriminatory or unintended outcomes.

Additionally, policies should support small and medium-sized enterprises (SMEs) in adopting ML technologies to bridge the digital divide with larger corporations. Governments can facilitate this by providing financial incentives, technical support, and training programs to help SMEs adopt ML technologies effectively. Investing in digital infrastructure, especially in underserved regions, is also essential to enable equitable access to advanced technologies and promote inclusive economic growth.

Public policy should also prioritize enhancing infrastructure, education, and workforce development. Integrating machine learning (ML) and artificial intelligence (AI) education into curricula and offering reskilling programs can prepare the labour

force for the evolving job market, thereby mitigating potential unemployment due to automation. Furthermore, establishing mechanisms to strengthen collaboration among academia, industry, and government is essential, as such synergies can accelerate the development of ML solutions that are ethically sound and aligned with real-world needs. Simultaneously, policymakers must address infrastructure challenges associated with ML implementation, including providing adequate computational resources and addressing data security and privacy concerns. Developing innovative approaches to ensure compliance with standards and guidelines for data management and cybersecurity is also important to safeguard individual rights and promote the responsible use of ML technologies.

Limitations and future work. In discussing the limitations of this study, it is important to acknowledge the recent mainstream emergence of generative AI-based applications. Our research reveals the intellectual structure and identifiable trends of machine learning (ML) and artificial intelligence (AI) applications in business and management as of August 2022, just before large language models (LLMs) and generative AI applications (e.g., ChatGPT) became prevalent. While our results indicate that these generative AI technologies were not yet significant in business and management contexts at that time, we recognize their rapidly expanding use cases since 2023.

This acceleration suggests that the current intellectual structure of ML and AI could become heavily skewed toward research focusing exclusively on generative AI, potentially overshadowing other established ML and AI algorithms and methods. Therefore, our study offers a "pre-generative AI boom" snapshot of proven, mature, and stable ML-based technologies and applications in business and management. It serves as a valuable reference for practitioners and academics to apply in their business cases and to guide future research and development.

For researchers aiming to replicate this work to capture an updated intellectual structure of ML and AI applications in business and management using a co-word analysis approach, it is crucial to be mindful of the timeless and non-historical nature of the clustering and graph algorithms involved. Without careful consideration, there is a risk of obtaining misleading results dominated by the surge of generative AI-focused publications from the past two years, which could obscure other historically significant ML and AI methods.

Given this context, we suggest that a balanced and realistic up-to-date intellectual structure of ML and AI research in business and management should distinguish between generative AI technologies and use cases and those not utilizing generative AI. This differentiation would lead to a two-part structure that provides a comprehensive taxonomy of ML and AI applications in business contexts, ensuring that both emerging and established technologies are appropriately represented.

Data availability

The data supporting the findings of this study are available in the Journal's Dataverse repository at https://doi.org/10.7910/DVN/YKAVDV.

Received: 4 April 2024; Accepted: 7 January 2025;
Published online: 10 March 2025

References

Adadi A, Berrada M (2018) Peeking inside the Black-Box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Ahmed NK, Atiya AF, Gayar NE, El-Shishiny H (2010) An empirical comparison of machine learning models for time series forecasting. Econ Rev 29:594–621. https://doi.org/10.1080/07474938.2010.481556

Atlantico (2023) The 2023 Latin America digital transformation report. Atlantico

Aziz S, Dowling M (2019) Chapter 3: Machine learning and AI for risk management. In Disrupting Finance: FinTech and Strategy in the 21st Century (pp. 33–50). Springer International Publishing. https://doi.org/10.1007/978-3-030-02330-0

Baek S, Mohanty SK, Glambosky M (2020) COVID-19 and stock market volatility: an industry level analysis. Financ Res Lett 37:101748. https://doi.org/10.1016/j.frl.2020.101748

Bakhshinategh B, Zaiane OR, ElAtia S, Ipperciel D (2018) Educational data mining applications and tasks: A survey of the last 10 years. Educ Inf Technol 23:537–553. https://doi.org/10.1007/s10639-017-9616-z

Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J et al. (2020) Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bekhet HA, Eletter SFK (2014) Credit risk assessment model for Jordanian commercial banks: neural scoring approach. Rev Dev Financ 4:20–28. https://doi.org/10.1016/j.rdf.2014.03.002

Bertolini M, Mezzogori D, Neroni M, Zammori F (2021) Machine Learning for industrial applications: a comprehensive literature review. Expert Syst Appl 175:114820. https://doi.org/10.1016/j.eswa.2021.114820

Bertsimas D, Kallus N (2020) From predictive to prescriptive analytics. Manag Sci 66:1025–1044. https://doi.org/10.1287/mnsc.2018.3253

Callon M, Courtial J-P, Laville F (1991) Co-word analysis as a tool for describing the network of interactions between basic and technological research: the case of polymer chemistry. Scientometrics 22:155–205. https://doi.org/10.1007/BF02019280

Canhoto AI, Clear F (2020) Artificial intelligence and machine learning as business tools: a framework for diagnosing value destruction potential. Bus Horiz 63:183–193. https://doi.org/10.1016/j.bushor.2019.11.003

Chang HW, Tai YC, Hsu JYJ (2010) Context-aware taxi demand hotspots prediction. IJBIDM 5:3. https://doi.org/10.1504/IJBIDM.2010.030296

Chen L-F, Tsai C-T (2016) Data mining framework based on rough set theory to improve location selection decisions: a case study of a restaurant chain. Tour Manag 53:197–206. https://doi.org/10.1016/j.tourman.2015.10.001

Chen M, Wu H (2005) An association-based clustering approach to order batching considering customer demand patterns. Omega 33:333–343. https://doi.org/10.1016/j.omega.2004.05.003

Di Vaio A, Palladino R, Hassan R, Escobar O (2020) Artificial intelligence and business models in the sustainable development goals perspective: a systematic literature review. J Bus Res 121:283–314. https://doi.org/10.1016/j.jbusres.2020.08.019

Dou Y, Liu Z, Sun L, Deng Y, Peng H, Yu PS (2020) Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 315–324. https://doi.org/10.1145/3340531.3411903

Dove G, Halskov K, Forlizzi J, Zimmerman J (2017) UX design innovation: Challenges for working with machine learning as a design material. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 278–288. https://doi.org/10.1145/3025453.3025739

Fernandes E, Holanda M, Victorino M et al. (2019) Educational data mining: predictive analysis of academic performance of public school students in the capital of Brazil. J Bus Res 94:335–343. https://doi.org/10.1016/j.jbusres.2018.02.012

Ferreira KJ, Lee BHA, Simchi-Levi D (2016) Analytics for an online retailer: demand forecasting and price optimization. M&SOM 18:69–88. https://doi.org/10.1287/msom.2015.0561

Fotiadis A, Polyzos S, Huan T-CTC (2021) The good, the bad and the ugly on COVID-19 tourism recovery. Ann Tour Res 87:103117. https://doi.org/10.1016/j.annals.2020.103117

Fuchs M, Höpken W, Lexhagen M (2014) Big data analytics for knowledge generation in tourism destinations—a case from Sweden. J Destin Mark Manag 3:198–209. https://doi.org/10.1016/j.jdmm.2014.08.002

Geiger A, Liu D, Alnegheimish S, Cuesta-Infante A, Veeramachaneni K (2020) Tadgan: Time series anomaly detection using generative adversarial networks. 2020 IEEE International Conference on Big Data (Big Data), 33–43. https://doi.org/10.1109/BigData50022.2020.9378139

Ghiassi M, Saidane H, Zimbra DK (2005) A dynamic artificial neural network model for forecasting time series events. Int J Forecast 21:341–362. https://doi.org/10.1016/j.ijforecast.2004.10.008

Goralski MA, Tan TK (2020) Artificial intelligence and sustainable development. Int J Manag Educ 18:100330. https://doi.org/10.1016/j.ijme.2019.100330

Graham B, Bonner K (2022) One size fits all? Using machine learning to study heterogeneity and dominance in the determinants of early-stage entrepreneurship. J Bus Res 152:42–59. https://doi.org/10.1016/j.jbusres.2022.07.043

Gray GL, Debreceny RS (2014) A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. Int J Account Inf Syst 15:357–380. https://doi.org/10.1016/j.accinf.2014.05.006

Grekousis G, Manetos P, Photis YN (2013) Modeling urban evolution using neural networks, fuzzy logic and GIS: the case of the Athens metropolitan area. Cities 30:193–203. https://doi.org/10.1016/j.cities.2012.03.006

Gu S, Kelly B, Xiu D (2020) Empirical asset pricing via machine learning. Rev Financ Stud 33:2223–2273. https://doi.org/10.1093/rfs/hhaa009

Guo Y, Barnes SJ, Jia Q (2017) Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. Tour Manag 59:467–483. https://doi.org/10.1016/j.tourman.2016.09.009

Hartmann J, Huppertz J, Schamp C, Heitmann M (2019) Comparing automated text classification methods. Int J Res Mark 36:20–38. https://doi.org/10.1016/j.ijresmar.2018.09.009

Hellas A, Ihantola P, Petersen A, Ajanovski VV, Gutica M, Hynninen T, Knutas A, Leinonen J, Messom C, Liao SN (2018) Predicting academic performance: A systematic literature review. Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education, 175–199. https://doi.org/10.1145/3293881.3295783

Hu K (2023) ChatGPT sets record for fastest-growing user base—analyst note. Reuters

Huang M-H, Rust RT (2021) A strategic framework for artificial intelligence in marketing. J Acad Mark Sci 49:30–50. https://doi.org/10.1007/s11747-020-00749-9

Huang Z, van der Aalst WMP, Lu X, Duan H (2011) Reinforcement learning based resource allocation in business process management. Data Knowl Eng 70:127–145. https://doi.org/10.1016/j.datak.2010.09.002

Jin X-H, Zhang G (2011) Modelling optimal risk allocation in PPP projects using artificial neural networks. Int J Proj Manag 29:591–603. https://doi.org/10.1016/j.ijproman.2010.07.011

Kaastra I, Boyd M (1996) Designing a neural network for forecasting financial and economic time series. Neurocomputing 10:215–236. https://doi.org/10.1016/0925-2312(95)00039-9

Kalinic Z, Marinkovic V, Molinillo S, Liébana-Cabanillas F (2019) A multi-analytical approach to peer-to-peer mobile payment acceptance prediction. J Retail Consum Serv 49:143–153. https://doi.org/10.1016/j.jretconser.2019.03.016

Kaplan A, Haenlein M (2020) Rulers of the world, unite! The challenges and opportunities of artificial intelligence. Bus Horiz 63:37–50. https://doi.org/10.1016/j.bushor.2019.09.003

Kemper RV, Roberts JM, Goodwin RD (1983) Tourism as a cultural domain the case of Taos, New Mexico. Ann Tour Res 10:149–171. https://doi.org/10.1016/0160-7383(83)90120-2

Khandani AE, Kim AJ, Lo AW (2010) Consumer credit-risk models via machine-learning algorithms. J Bank Financ 34:2767–2787. https://doi.org/10.1016/j.jbankfin.2010.06.001

Koutanaei FN, Sajedi H, Khanbabaei M (2015) A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. J Retail Consum Serv 27:11–23. https://doi.org/10.1016/j.jretconser.2015.07.003

Kraus M, Feuerriegel S, Oztekin A (2020) Deep learning in business analytics and operations research: models, applications and managerial implications. Eur J Oper Res 281:628–641. https://doi.org/10.1016/j.ejor.2019.09.018

Larson JS, Bradlow ET, Fader PS (2005) An exploratory look at supermarket shopping paths. Int J Res Mark 22:395–414. https://doi.org/10.1016/j.ijresmar.2005.09.005

Law R, Li G, Fong DKC, Han X (2019) Tourism demand forecasting: a deep learning approach. Ann Tour Res 75:410–423. https://doi.org/10.1016/j.annals.2019.01.014

Lee D, Hosanagar K, Nair HS (2018) Advertising content and consumer engagement on social media: evidence from Facebook. Manag Sci 64:5105–5131. https://doi.org/10.1287/mnsc.2017.2902

Luo X, Wu C, Rosenberg D, Barnes D (2009) Supplier selection in agile supply chains: an information-processing model and an illustration. J Purch Supply Manag 15:249–262. https://doi.org/10.1016/j.pursup.2009.05.004

Ma Y, Xiang Z, Du Q, Fan W (2018) Effects of user-provided photos on hotel review helpfulness: an analytical approach with deep leaning. Int J Hosp Manag 71:120–131. https://doi.org/10.1016/j.ijhm.2017.12.008

Manela A, Moreira A (2017) News implied volatility and disaster concerns. J Financ Econ 123:137–162. https://doi.org/10.1016/j.jfineco.2016.01.032

McCarty JA, Hastak M (2007) Segmentation approaches in data-mining: a comparison of RFM, CHAID, and logistic regression. J Bus Res 60:656–662. https://doi.org/10.1016/j.jbusres.2006.06.015

Menon S, Yadav J, Chopra A, Thomas J (2024) Strategic integration of analytics and artificial intelligence in sustainable human resource management: Fostering HR excellence. 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 1–5. https://doi.org/10.1109/ICRITO61523.2024.10522141

Miklosik A, Krah AB (2023) Pinpointing the driving forces propelling digital business transformation. J Risk Financ Manag 16:488. https://doi.org/10.3390/jrfm16110488

Miller T (2019) Explanation in artificial intelligence: insights from the social sciences. Artif Intell 267:1–38. https://doi.org/10.1016/j.artint.2018.07.007

Morency C, Trépanier M, Agard B (2007) Measuring transit use variability with smart-card data. Transp Policy 14:193–203. https://doi.org/10.1016/j.tranpol.2007.01.001

Mullainathan S, Spiess J (2017) Machine learning: an applied econometric approach. J Econ Perspect 31:87–106. https://doi.org/10.1257/jep.31.2.87

Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69:026113. https://doi.org/10.1103/PhysRevE.69.026113

Ngai EWT, Wu Y (2022) Machine learning in marketing: a literature review, conceptual framework, and research agenda. J Bus Res 145:35–48. https://doi.org/10.1016/j.jbusres.2022.02.049

Noack A (2007) Energy models for graph clustering. JGAA 11:453–480. https://doi.org/10.7155/jgaa.00154

Pai P-F, Lin C-S (2005) A hybrid ARIMA and support vector machines model in stock price forecasting. Omega 33:497–505. https://doi.org/10.1016/j.omega.2004.07.024

Pallathadka H, Mustafa M, Sanchez DT et al. (2023) Impact of machine learning on management, healthcare and agriculture. Mater Today: Proc 80:2803–2806. https://doi.org/10.1016/j.matpr.2021.07.042

Paschek D, Luminosu CT, Draghici A (2017) Automated business process management—in times of digital transformation using machine learning or artificial intelligence. MATEC Web Conf 121:04007. https://doi.org/10.1051/matecconf/201712104007

Paschen J, Wilson M, Ferreira JJ (2020a) Collaborative intelligence: how human and artificial intelligence create value along the B2B sales funnel. Bus Horiz 63:403–414. https://doi.org/10.1016/j.bushor.2020.01.003

Paschen U, Pitt C, Kietzmann J (2020b) Artificial intelligence: building blocks and an innovation typology. Bus Horiz 63:147–155. https://doi.org/10.1016/j.bushor.2019.10.004

Preece A (2018) Asking 'Why' in AI: explainability of intelligent systems—perspectives and challenges. Intell Syst Account Financ Manag 25:63–72. https://doi.org/10.1002/isaf.1422

Rojas-Córdova C, Heredia-Rojas B, Ramírez-Correa P (2020) Predicting business innovation intention based on perceived barriers: a machine learning approach. Symmetry 12:1381. https://doi.org/10.3390/sym12091381

Samoilenko S, Osei-Bryson K-M (2013) Using data envelopment analysis (DEA) for monitoring efficiency-based performance of productivity-driven organizations: design and implementation of a decision support system. Omega 41:131–142. https://doi.org/10.1016/j.omega.2011.02.010

Sedera D, Tan C-W, Xu D (2022) Digital business transformation in innovation and entrepreneurship. Inf Manag 59:103620. https://doi.org/10.1016/j.im.2022.103620

Singh VK, Singh P, Karmakar M et al. (2021) The journal coverage of Web of Science, Scopus and Dimensions: a comparative analysis. Scientometrics 126:5113–5142. https://doi.org/10.1007/s11192-021-03948-5

Stanford "AI Index Report" (2023) THE AI Index Report Measuring trends in artificial intelligence. Stanford Institute for Human-Centered Artificial Intelligence

Sullivan R, Timmermann A, White H (2001) Dangers of data mining: the case of calendar effects in stock returns. J Econom 105:249–286. https://doi.org/10.1016/S0304-4076(01)00077-X

Syam N, Sharma A (2018) Waiting for a sales renaissance in the fourth industrial revolution: machine learning and artificial intelligence in sales research and practice. Ind Mark Manag 69:135–146. https://doi.org/10.1016/j.indmarman.2017.12.019

Tabesh P, Mousavidin E, Hasani S (2019) Implementing big data strategies: a managerial perspective. Bus Horiz 62:347–358. https://doi.org/10.1016/j.bushor.2019.02.001

Thiprungsri S, Vasarhelyi M (2011) Cluster analysis for anomaly detection in accounting data: an audit approach. IJDAR 11. https://doi.org/10.4192/1577-8517-v11_4

Urquhart A (2017) Price clustering in Bitcoin. Econ Lett 159:145–148. https://doi.org/10.1016/j.econlet.2017.07.035

van Eck NJ, Waltman L (2007) VOS: a new method for visualizing similarities between objects. In: Decker R, Lenz H-J (eds) Advances in data analysis. Springer, Berlin, Heidelberg, pp. 299–306

Vial G (2019) Understanding digital transformation: a review and a research agenda. J Strateg Inf Syst 28:118–144. https://doi.org/10.1016/j.jsis.2019.01.003

Vidya CT, Prabheesh KP (2020) Implications of COVID-19 pandemic on the global trade networks. Emerg Mark Financ Trade 56:2408–2421. https://doi.org/10.1080/1540496X.2020.1785426

Waltman L, van Eck NJ, Noyons ECM (2010) A unified approach to mapping and clustering of bibliometric networks. J Informetr 4:629–635. https://doi.org/10.1016/j.joi.2010.07.002

Wei S, Ye N, Zhang S, Huang X, Zhu J (2012) Collaborative filtering recommendation algorithm based on item clustering and global similarity. 2012 Fifth International Conference on Business Intelligence and Financial Engineering, 69–72. https://doi.org/10.1109/BIFE.2012.23

Wenzel H, Smit D, Sardesai S (2019) A literature review on machine learning in supply chain management. https://doi.org/10.15480/882.2478

Xiang Z, Du Q, Ma Y, Fan W (2017) A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. Tour Manag 58:51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Zaki M (2019) Digital transformation: harnessing digital technologies for the next generation of services. JSM 33:429–435. https://doi.org/10.1108/JSM-01-2019-0034

Zhang JZ, Srivastava PR, Sharma D, Eachempati P (2021) Big data analytics and machine learning: a retrospective overview and bibliometric analysis. Expert Syst Appl 184:115561. https://doi.org/10.1016/j.eswa.2021.115561

Zhang K, Chen Y, Li C (2019) Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: the case of Beijing. Tour Manag 75:595–608. https://doi.org/10.1016/j.tourman.2019.07.002

Zhang W, Wang P, Li X, Shen D (2018) Some stylized facts of the cryptocurrency market. Appl Econ 50:5950–5965. https://doi.org/10.1080/00036846.2018.1488076

Zhao F, Shi B, Liu R et al. (2018) Theme trends and knowledge structure on choroidal neovascularization: a quantitative and co-word analysis. BMC Ophthalmol 18:86. https://doi.org/10.1186/s12886-018-0752-z

Zhou Z-H (2021) Machine learning. Springer, Singapore, Singapore

Author contributions

Aglaya Batz (AB): Designed and developed the analytical framework, structured the methodological approach and contributed to the cluster analysis, interpretation of the results, and discussion and conclusion sections. AB also revised and drafted the final version of the manuscript and approved it for submission, ensuring the accuracy and completeness of all aspects of the work. David F. D'Croz-Barón (DFDB): Interpreted the results and provided expertise in machine learning (ML) and artificial intelligence (AI) from an industry perspective. DFDB also identified potential future developments in the trajectory of clusters and critically reviewed the manuscript. Carlos Jesús Vega Perez (CJVP): Interpreted the results and provided expertise in machine learning (ML) and artificial intelligence (AI) from an industry perspective. CJVP also identified potential future developments in the trajectory of clusters and critically reviewed the manuscript. Carlos A. Ojeda-Sanchez (CAOS): Executed the methodological procedures and wrote the initial draft of the manuscript. CAOS also contributed to the analysis of results and the discussion section.

Competing interests

The authors declare no competing interests.

Ethical approval

Ethical approval was not required as this study did not involve human participants.

Informed consent

Informed consent was not required as this study did not involve human participants.

Additional information

Correspondence and requests for materials should be addressed to Aglaya Batz.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2025