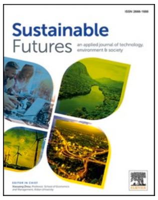
Sustainable Futures 9 (2025) 100423

Contents lists available at ScienceDirect

Sustainable Futures

journal homepage: www.sciencedirect.com/journal/sustainable-futures



Examining the synergies between industry 4.0 and sustainability dimensions using text mining, sentiment analysis, and association rules

Mohamad Ali Saleh Saleh [a,*], Mutaz AlShafeey [b]

[a] Institute of Social Sciences, Department of Economics and Management, University of Dunaújváros, Dunaújváros, Táncsics Mihály u. 1/a, 2400, Hungary
[b] Institute of Data Analytics and Information Systems, Corvinus University of Budapest, Budapest, Fővám tér 13-15, H-1093, Hungary

ARTICLE INFO

Keywords:
Industry 4.0
Sustainability
Sustainable development
Economic dimension
Social dimension
Environmental dimension

ABSTRACT

The transformation to Industry 4.0 has significantly revolutionized manufacturing and production processes, raising important questions about their impact on sustainability. This study aims to explore the interplay between Industry 4.0 and the economic, social, and environmental dimensions of sustainability. The methodological approach includes advanced text-mining, sentiment analysis, and association rule-mining techniques to examine 6,759 abstracts from the Scopus database. The text mining highlighted frequent keywords related to Industry 4.0 and the three sustainability dimensions, characterized by "economic growth," "circular economy," "social responsibility," "education 4.0," "energy efficiency," and "waste management." Sentiment analysis revealed a predominantly positive perspective, with 2,608 positive sentiments out of 2,761 in the economic dimension, 1,604 out of 1,728 in the social dimension, and 1,352 out of 1,527 in the environmental dimension. The association rule mining uncovered the associations between Industry 4.0 and each sustainability dimension. The highest support was observed between Industry 4.0 and economic sustainability, with a support value of 0.444, confidence of 0.855, and a lift of 1.060. These findings highlight the role of Industry 4.0 in promoting resource efficiency and reducing waste through circular economy principles and advanced manufacturing technologies. For the social dimension, the analysis revealed a strong association with Industry 4.0 (support: 0.430, confidence: 0.831, lift: 1.030), emphasizing its role in enhancing worker safety and job satisfaction by automating hazardous tasks and creating new high-tech job opportunities. In the environmental dimension, a significant association was found (support: 0.380, confidence: 0.827, lift: 1.024), showing Industry 4.0′s contribution to sustainability through optimized energy consumption and emissions reduction as the integration of big data and IoT enables real-time monitoring of environmental impacts. The rule combining economic and social aspects with Industry 4.0 (support: 0.219, confidence: 0.87, lift: 1.078) highlights the interconnected nature of these dimensions, suggesting many studies consider economic and social dimensions together in the Industry 4.0 context.

1. Introduction

In recent years, both Industry 4.0 and sustainability have become increasingly prominent topics of discussion and analysis across academic, business management, and policy-making circles. Although Industry 4.0 and sustainability are widely recognized as important subjects, the nature of their relationship remains ambiguous, despite numerous studies attempting to elucidate their connection [1]. The concept of Industry 4.0, while widely discussed, lacks a universally agreed-upon definition in the academic and professional communities [2,3]. However, several scholars and institutions [3–6] define Industry 4.0 as a system encompassing a diverse array of cutting-edge technologies, systems, and methodologies, Castelo-Branco et al. [7] offer a more process-oriented definition. They characterize Industry 4.0 as an interconnected system of technologies, devices, and processes that operate cohesively across various production stages and multiple supply chain tiers. According to Castelo-Branco et al. [7] the key features of this system include enabling autonomous production, seamlessly integrated operations, decentralized decision-making, and minimal human involvement. This comprehensive approach incorporates both the technological components and their systemic integration, encompassing the Internet of Things (IoT), Cyber-Physical Systems (CPS), autonomous

---

* Corresponding author.
E-mail address: mohamad.saleh@uniduna.hu (M.A.S. Saleh).

https://doi.org/10.1016/j.sftr.2024.100423
Received 25 July 2024; Received in revised form 17 November 2024; Accepted 18 December 2024

robotics, immersive visualization tools (such as virtual and augmented reality), cloud computing infrastructure, blockchain technology, advanced data analytics, additive manufacturing processes, and digital twin simulations. These interconnected technologies form the backbone of Industry 4.0, enabling the seamless flow of information and control across the entire production and supply chain system.

The emergence and evolution of Industry 4.0 intersects with broader discussions surrounding sustainable industrial practices and development. Sustainability has been subject to various interpretations and definitions as a comprehensive and multifaceted concept. Among these, the most widely cited originates from the Brundtland report, which introduced the term “Sustainable development”. UN World Commission on Environment and Development [8] definition describes sustainable development as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs”. As the concept of sustainability has evolved, it has come to be understood as the pursuit of equilibrium among economic, social, and environmental outcomes a framework known as the triple bottom line (TBL). In the business world, this translates to recognizing that an organization’s enduring success is contingent upon its performance across all three sustainability dimensions [9].

Contemporary research has attempted to explore the relationship between Industry 4.0 and sustainability. These investigations have revealed a complex interplay, characterized by both synergies and potential conflicts. From an economic perspective, Industry 4.0 technologies can enhance process optimization, leading to more efficient resource utilization and potentially reducing operational costs [10,11]. These advancements can also support sustainable business models such as the circular economy [12]. Regarding the social dimension, Industry 4.0′s impact on sustainability performance is multifaceted. It influences various aspects of work, including working conditions, hours, required skills, and occupational health and safety. Research has revealed both positive and negative correlations in this area, suggesting a complex relationship between technological advancement and social sustainability [13–16]. In terms of environmental sustainability, Industry 4.0 presents both opportunities and challenges. On one hand, it can contribute positively by reducing waste production [13,16] and enabling more effective energy monitoring [17,18]. It also facilitates eco-friendly practices like green cloud computing. However, the proliferation of Industry 4.0 technologies may also lead to increased electronic waste and higher energy consumption [19], highlighting the need for careful management of these innovations.

The current study aims to explore the dominant themes, prevailing sentiments, and intricate relationships between Industry 4.0 and sustainability. It employs text mining, including bigram analysis, to identify and quantify key concepts and their frequencies within the abstracts. Sentiment analysis is used to reveal the overall attitude (positive, negative, or neutral) towards Industry 4.0′s impact on each sustainability dimension. Text mining refers to the process of extracting meaningful insights and patterns from large volumes of unstructured text data, such as research abstracts. It often involves techniques like keyword analysis and bigram analysis. Bigram analysis is a text mining technique that identifies frequent two-word combinations within the text. This helps reveal common themes and concepts that are prevalent in the analyzed corpus. These techniques allow researchers to systematically analyze a large amount of textual data to uncover the key focus areas, trends, and relationships that may not be easily detectable through manual review. This provides data-driven insights into the intersections between Industry 4.0 and sustainability. Whereas, Sentiment analysis evaluates the overall attitudes (positive, negative, or neutral) expressed towards a particular topic or concept within the analyzed text. It provides quantitative measures of the prevailing sentiment. Sentiment analysis gives a more detailed understanding of how Industry 4.0 technologies are perceived in terms of their impact on the different dimensions of sustainability. The prevalence of positive, negative, or neutral sentiments can indicate the overall perspective and expectations surrounding these technologies. Association rule mining is utilized to uncover significant relationships and the probability of co-occurrence between Industry 4.0 and various sustainability dimensions. The study objectives include identifying the primary focus areas within each sustainability dimension, assessing the distribution of research attention, evaluating the prevailing sentiment towards Industry 4.0′s impact, and discovering meaningful associations between Industry 4.0 and specific sustainability aspects.

This study offers several novel contributions and new discoveries compared to previous studies, as it adopts a holistic approach by simultaneously addressing the environmental, economic, and social dimensions of sustainability, providing a more integrated understanding of how Industry 4.0 influences sustainable development. The study employs a combination of advanced machine learning methods, including text mining, sentiment analysis, and association rule mining, which allows for a more detailed analysis of the interplay between Industry 4.0 and sustainability dimensions. The study analyzes 6759 abstracts retrieved from Scopus, leveraging a large corpus of academic content to enhance the robustness and generalizability of the findings.

The current study employed text mining, including bigram analysis, to identify key themes and concepts related to Industry 4.0 and sustainability. Frequent keywords identified include “economic growth,” “circular economy,” “social responsibility,” “education 4.0,” “energy efficiency,” and “waste management,” emphasizing the intersections between Industry 4.0 and various sustainability dimensions. Sentiment analysis indicated a predominantly positive perspective on Industry 4.0′s impact, with high positive sentiment scores across the economic (94.5 %), social (92.8 %), and environmental (88.5 %) dimensions. This reflects the perceived potential of Industry 4.0 technologies to drive sustainable development. Association rule mining uncovered strong relations between Industry 4.0 and each sustainability dimension. The association between economic dimension highlights the significant role of Industry 4.0 in promoting resource efficiency and reducing waste, particularly through the adoption of circular economy principles and advanced manufacturing technologies like additive manufacturing. In the social dimension, Industry 4.0 technologies enhance worker safety and job satisfaction by automating hazardous tasks and creating new job opportunities in high-tech sectors. In the environmental dimension, Industry 4.0 contributes to sustainability by optimizing energy consumption and reducing emissions through smart manufacturing practices. The integration of big data and IoT allows for real-time monitoring and control of environmental impacts, facilitating sustainable manufacturing practices. The analysis indicates a predominantly positive perspective towards Industry 4.0, reflecting the perceived potential of these technologies to drive sustainable development across economic, social, and environmental dimensions.

In comparison to previous studies, such as [20,1], and [21], this study provides a more detailed quantitative analysis and empirical examination. The compared studies are primarily qualitative, focusing on reviews, conceptualizations, and bibliometric analyses. In contrast, this study combines both qualitative and quantitative approaches. The qualitative aspect is addressed through text mining, bigram analysis, and sentiment analysis, while the quantitative aspect is achieved through association rule mining. The co-occurrence of keywords facilitates this important transfer of qualitative data into quantitative data. Beltrami et al. [1] used conceptualization and theorization to explore the relationship between Industry 4.0 and sustainability, offering valuable theoretical insights but lacking empirical analysis. Ejsmont et al. [21] conducted a bibliometric literature review to assess the impact of Industry 4.0 on sustainability, providing a broad overview of the research landscape but not exploring the specific associations and sentiments uncovered by the current study’s advanced analytical methods. This study advances the understanding of the interplay between Industry 4.0 and sustainability by providing a comprehensive, data-driven analysis that highlights the associations and sentiments towards Industry 4.0 the economic, social, and environmental dimensions of sustainability.

2. Literature

Despite the growing body of research on Industry 4.0 and sustainability, the relationship between these two concepts remains ambiguous for several reasons. One significant factor contributing to this ambiguity is the lack of universally agreed-upon definitions for both Industry 4.0 and sustainability. Industry 4.0 encompasses a wide range of technologies and methodologies, leading to varied interpretations and applications across different studies [2,3]. Similarly, sustainability is a multifaceted concept that includes environmental, economic, and social dimensions, each with its own set of metrics and goals [22]. This diversity in definitions and scope complicates the understanding of how these two concepts interact. The interdisciplinary nature of the relationship between Industry 4.0 and sustainability further adds to the complexity. This relationship spans multiple disciplines, including engineering, economics, environmental science, and social sciences, making it challenging to develop a cohesive framework that captures all relevant aspects and their interactions [13,16]. Additionally, the rapid evolution of Industry 4.0 technologies and their emerging applications in sustainability mean that their impacts are not yet fully understood or documented, leading to gaps in the literature [23]. Moreover, existing research presents mixed findings on the impact of Industry 4.0 on sustainability. Some studies highlight positive outcomes, such as improved resource efficiency and reduced emissions [24], while others point to potential negative effects, such as increased electronic waste and higher energy consumption [19]. These conflicting results make it difficult to draw definitive conclusions about the relationship between Industry 4.0 and sustainability.

The ambiguity in the relationship between Industry 4.0 and sustainability can be seen as both a question of academic value and an academic frontier. On one hand, the lack of clear, consistent findings highlights the need for more thorough, comprehensive studies that can provide definitive insights. On the other hand, this ambiguity represents an academic frontier, offering opportunities for innovative research that can bridge existing gaps and advance our understanding of these complex interactions. From the perspective of this study, the specific manifestations of ambiguity include inconsistent metrics and indicators, varied methodological approaches, and context-dependent outcomes. Different studies use varying metrics and indicators to assess the impact of Industry 4.0 on sustainability, making it difficult to compare results and draw general conclusions. The use of diverse methodological approaches, ranging from qualitative case studies to quantitative modeling, leads to different interpretations and findings. Additionally, the impact of Industry 4.0 on sustainability can vary significantly depending on the specific context, such as the industry sector, geographic region, and stage of technological adoption. By addressing these reasons and manifestations of ambiguity, this study aims to provide a more integrated and comprehensive understanding of the interplay between Industry 4.0 and sustainability. Leveraging advanced analytical methods, this research seeks to uncover new insights and contribute to the ongoing academic discourse, ultimately helping to clarify the complex relationship between these two crucial areas.

Industry 4.0, often referred to as the Fourth Industrial Revolution, represents a transformative shift in manufacturing and production processes driven by advancements in digital technology. Originating in Germany in 2011, the term was introduced to enhance the country's manufacturing competitiveness through a strategic plan [25]. This concept has since evolved, encompassing a variety of technologies and methodologies aimed at creating "smart" factories. However, the term Industry 4.0 lacks a universally accepted definition, contributing to ambiguity in both academic and practical applications [3]. It overlaps with related concepts such as "smart manufacturing," "digital transformation," and the "industrial internet," adding to the confusion [26,27]. Industry 4.0 integrates technologies like the Internet of Things (IoT), cyber-physical systems (CPS), big data analytics, cloud computing, and artificial intelligence (AI), all aiming to create interconnected and intelligent manufacturing systems [2,28].

The concept has inspired similar initiatives globally, such as the US Advanced Manufacturing Partnership and the European Factories of the Future Program, reflecting a broader trend toward digital transformation in manufacturing [29]. The adoption of Industry 4.0 technologies promises a manufacturing renaissance by significantly enhancing productivity, flexibility, and resource efficiency [30,31]. Research on Industry 4.0 encompasses various dimensions including technological advancements, organizational changes, and impacts on labor and society. A systematic literature review by Culot et al. [3] categorized the key definitional elements of Industry 4.0, identifying common themes and contradictions among different definitions. Culot et al. [3] review also highlighted the necessity for clearer conceptual frameworks to guide both academic research and practical implementation.

Key technological drivers of Industry 4.0 include the Internet of Things (IoT), which facilitates real-time data exchange and monitoring across devices and systems; Cyber-Physical Systems (CPS), which integrate physical processes with computational models to enhance automation and control; Big Data Analytics, which enable the processing and analysis of large datasets to drive decision-making and predictive maintenance; Cloud Computing, which provides scalable and on-demand computing resources to support extensive data storage and processing needs; and Artificial Intelligence (AI), which enhances the ability to perform complex tasks, including machine learning and advanced robotics. The integration of Industry 4.0 technologies can significantly impact sustainability by optimizing resource use, reducing waste, and enhancing energy efficiency [32]. The environmental, economic, and social dimensions of sustainability are simultaneously addressed, promoting a holistic approach to sustainable development [33].

Culot et al. [3] propose several future research directions, emphasizing the need for comprehensive frameworks to better define and measure the impact of Industry 4.0, interdisciplinary studies combining technical, managerial, and social perspectives, and practical applications through case studies and real-world implementations to validate theoretical models. Industry 4.0 represents a significant shift in manufacturing, driven by digital technologies and aimed at creating more intelligent and efficient production systems. Despite definitional ambiguities, the concept has garnered global interest and initiated numerous research and practical initiatives. Ongoing research is crucial to fully understand and harness the potential of Industry 4.0, particularly in enhancing sustainability and driving future industrial advancements.

The concept of sustainability is frequently represented through the "three pillars" framework, encompassing environmental, economic, and social dimensions. This model aims to balance the trade-offs and synergies between these three areas, creating a comprehensive approach to sustainable development. The environmental pillar emphasizes the protection and preservation of natural resources and ecosystems. It involves efforts to reduce pollution, conserve biodiversity, and manage resources sustainably to ensure that the needs of the present do not compromise the ability of future generations to meet their own needs [22]. This pillar highlights the importance of maintaining the integrity of the planet's ecological systems, recognizing that economic and social well-being ultimately depend on a healthy environment [22]. According to Sachs [34], addressing climate change and other environmental challenges is central to achieving sustainability. The economic pillar focuses on promoting sustainable economic growth that provides prosperity and opportunities for all. It promotes economic practices that are not only profitable but also equitable and environmentally responsible. This involves encouraging innovations and efficiencies that reduce environmental impact, supporting industries and jobs that are sustainable, and ensuring that economic benefits are distributed fairly across society [22]. Sustainable economic practices aim to create long-term stability rather than short-term gains. The integration of sustainability into business practices is essential for long-term economic success [9]. The social pillar addresses issues related to human well-being, equity, and social justice. It seeks to ensure that all individuals have access to basic resources such as education, healthcare, and clean water and that communities are resilient and inclusive [22]. This pillar promotes social cohesion and cultural diversity and strives to protect human rights and improve the quality of life for all. By focusing on social sustainability, the goal is to build societies that are just, equitable, and capable of sustaining themselves over the long term. Social sustainability is fundamental for ensuring that the benefits of development reach all segments of society [35].

While the three pillars of sustainability are often presented as distinct categories, they are deeply interconnected. Effective sustainability strategies recognize the interdependencies between environmental health, economic vitality, and social well-being [22]. For example, a clean and healthy environment contributes to human health and economic productivity, just as economic development can provide the resources needed for environmental protection and social programs. Therefore, a holistic approach that integrates these three pillars is essential for achieving true sustainability. The integration of these pillars is critical to addressing global challenges and ensuring long-term sustainability [8]. The three pillars of sustainability, environmental, economic, and social, provide a framework for understanding and addressing the complex challenges of sustainable development. By balancing these interconnected dimensions, it is possible to create a sustainable future that benefits all aspects of society [8,9,22,34,35].

To effectively analyze the complex relationship between Industry 4.0 and sustainability in the existing literature, researchers have increasingly turned to advanced analytical methods. Text mining and Natural Language Processing (NLP) techniques have emerged as powerful tools for analyzing large volumes of unstructured text data in academic research. According to Kumar and Ravi [36], text mining encompasses various computational methods that enable researchers to discover patterns, extract meaningful information, and identify relationships within textual content through automated processes. This approach has become particularly valuable in analyzing scientific literature, where the volume of published research makes manual analysis impractical [37]. Bigram analysis, a specific text mining technique, focuses on identifying frequently co-occurring word pairs within text, helping researchers understand common themes and contextual relationships in the analyzed corpus [38]. As demonstrated by Hassani et al. [39], this approach has been successfully applied in various fields to uncover hidden patterns and trends in large text collections. The importance of bigram analysis lies in its ability to preserve contextual meaning that might be lost in single-word (unigram) analysis, making it particularly useful for understanding complex topics and their relationships [40].

Sentiment analysis, also known as opinion mining, is a computational method that determines the emotional tone or attitude expressed in text [41,42] define sentiment analysis as the computational treatment of opinions, sentiments, and subjectivity in text. In academic research, sentiment analysis helps quantify subjective information and understand attitudes towards specific topics or concepts, providing valuable insights into how different subjects are perceived in the literature [43]. The application of sentiment analysis has grown significantly in recent years, with researchers using it to analyze everything from social media content to scientific abstracts [44].

Using these analytical approaches, researchers have been able to systematically examine how Industry 4.0, characterized by the integration of advanced digital technologies such as the Internet of Things (IoT), cyber-physical systems, big data analytics, and artificial intelligence (AI), significantly impacts economic, social, and environmental sustainability [23].

The relationship between Industry 4.0 and sustainability is expressed across economic, social, and environmental dimensions, with each influencing the others in unique and interconnected ways. The economic dimension of Industry 4.0′s relationship with sustainability has been extensively studied, revealing multiple pathways through which digital technologies enhance economic sustainability. Lee et al. [45] provide evidence of how predictive maintenance systems reduce operational costs while improving worker safety and environmental performance by preventing equipment failures and reducing emergency maintenance situations. Kamble et al. [46] further demonstrate how smart manufacturing practices, including predictive maintenance and real-time monitoring systems, significantly improve operational efficiency and reduce costs. Their research shows that Industry 4.0 technologies can reduce maintenance costs by up to 20 % while improving overall equipment effectiveness. Frank et al. [47] elaborate on how Industry 4.0 drives economic sustainability through enhanced productivity and innovation, finding that companies implementing Industry 4.0 technologies report 15–30 % improvements in inventory management and up to 25 % increases in labor productivity. The economic benefits extend beyond direct operational improvements to enable new business models. Bressanelli et al. [12] document how digital technologies enable product-service systems and sharing economy models that enhance both economic and environmental sustainability. This is supported by Nascimento et al. [48], who highlight how Industry 4.0 technologies facilitate the transition to circular economy models, creating new revenue streams through resource optimization and waste reduction.

The transformation of supply chain management through Industry 4.0 technologies represents another crucial economic dimension. Saberi et al. [49] highlight how blockchain technology and IoT integration enhance supply chain transparency and traceability, reducing transaction costs and improving coordination among supply chain partners. Bag et al. [50] further demonstrate how advanced analytics and artificial intelligence enable more accurate demand forecasting and inventory optimization, leading to significant cost reductions and improved resource efficiency. Their research indicates that digitally integrated supply chains can achieve up to a 30 % reduction in operational costs. These technological advances not only improve economic performance but also enable better monitoring and verification of sustainable practices throughout the supply chain.

The social dimension of Industry 4.0 presents both opportunities and challenges for sustainable development. Frey and Osborne [51] estimate that while up to 47 % of current jobs might be at risk of automation, Industry 4.0 creates new high-skilled positions in technology-related fields. Benešová and Tupa [52] emphasize the critical role of education and training in this transition, asserting that successful implementation of Industry 4.0 requires comprehensive workforce development programs to equip workers with necessary digital skills. Fantini et al. [53] examine how automation and robotics enhance workplace safety by taking over hazardous tasks, while simultaneously creating concerns about job displacement. Their research indicates that while certain routine jobs may be automated, new roles emerge in areas such as robotics maintenance, data analytics, and system integration.

The social implications of Industry 4.0 extend beyond employment to broader aspects of workplace quality and social inclusion. Ghobakhloo [23] describes how digital technologies improve working conditions through enhanced human-machine interfaces and more flexible work arrangements. This research shows that Industry 4.0 technologies can reduce workplace accidents by up to 25 % while improving job satisfaction through the elimination of repetitive tasks. Rauch et al. [54] explore how digital transformation affects organizational culture and worker engagement, finding that companies successfully implementing Industry 4.0 technologies report higher levels of employee satisfaction and improved work-life balance. However, they also note the importance of managing the digital divide and ensuring inclusive access to technology-enabled opportunities.

In terms of environmental sustainability, Industry 4.0 technologies enable substantial improvements in environmental performance through various mechanisms. Rong et al. [17] demonstrate how smart grid integration and energy monitoring systems, enabled by Industry 4.0 technologies, facilitate better integration of renewable energy sources and more efficient energy utilization across industrial processes. Ford and Despeisse [19] elaborate on how additive manufacturing technologies, a key component of Industry 4.0, support environmental sustainability by minimizing material waste during production and enabling the creation of lightweight components that improve overall product efficiency. Stock et al. [13] show how Industry 4.0 technologies enable significant improvements in environmental performance through advanced monitoring, control, and optimization of industrial processes. Their research shows that smart manufacturing systems can reduce energy consumption by up to 30 % through real-time optimization and predictive maintenance. Bai et al. [24] further describes how Industry 4.0 technologies, particularly IoT sensors and advanced analytics, enable more precise control of energy usage and emissions in manufacturing processes, leading to significant reductions in environmental impact.

The integration of Industry 4.0 technologies with circular economy principles has emerged as a powerful driver of environmental sustainability. Rosa et al. [55] examine how digital twins and advanced simulation models enable companies to design products specifically for recyclability and remanufacturing, extending product lifecycles and reducing waste. Their research indicates that companies implementing these technologies can achieve up to a 25 % reduction in material waste and a 20 % improvement in resource efficiency. Energy management and efficiency represent another crucial environmental benefit of Industry 4.0 implementation. Ghobakhloo [23] documents how smart manufacturing systems utilize IoT devices and advanced analytics to optimize energy consumption patterns in real time. Their research shows that intelligent energy management systems can reduce industrial energy consumption by 20–30 % while simultaneously improving production efficiency.

The integration of economic, social, and environmental dimensions in Industry 4.0 creates significant synergies for sustainable development. de Sousa Jabbour et al. [56] suggest that successful implementation of Industry 4.0 requires consideration of all three sustainability dimensions to ensure long-term viability. Stock et al. [13] further demonstrate how digital technologies simultaneously improve economic performance, social outcomes, and environmental metrics through enhanced transparency, better working conditions, and more efficient resource use. Cui et al. [57] highlight how real-time monitoring and control systems enabled by Industry 4.0 technologies create safer working environments while optimizing resource efficiency and reducing environmental impact. These studies collectively suggest that successful digital transformation requires attention to all three dimensions to maximize sustainable development outcomes.

3. Methodology

3.1. Research design and approaches

This study aims to examine the influence of Industry 4.0 on the environmental, economic, and social dimensions of sustainability through three methods: text mining, sentiment analysis, and association rule mining as shown in Fig. 1. The initial step involved retrieving 6759 abstracts from Scopus.com. The inclusion and exclusion criteria for selecting samples were defined as follows: The search was conducted using the keywords “industry 4.0,” “environmental,” “economic,” and “social dimensions of sustainability,” to ensure the database contained relevant articles addressing the core topic. By focusing on the targeted set of keywords, the study employs an unsupervised learning approach to uncover emergent themes within the sustainability pillars related to Industry 4.0, rather than imposing a predetermined list. This method allows patterns to emerge naturally from the data, avoiding bias from pre-existing categorizations. Expanding the keyword set would increase data complexity and reduce the precision of the analysis, making it less manageable and efficient. Scopus was chosen as the primary database for this study because of its high-quality, curated content and structured indexing, which provide reliable metadata and ensure robust data quality. In contrast, Google Scholar, while offering broader coverage, includes non-peer-reviewed content and lacks the rigorous quality controls necessary for consistent analysis. Additionally, incorporating multiple databases would demand additional resources and face constraints from API and subscription limitations, potentially complicating the analysis and impacting its precision.

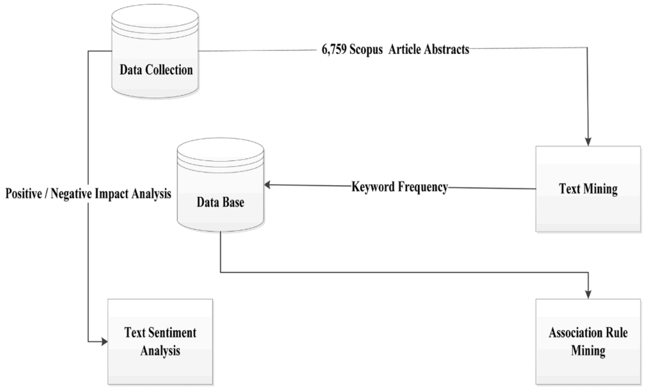


Fig. 1. Flow chart of the methods utilized.
Source: Authors.

The documents types included in the search were articles, conference papers, book chapters, reviews, and conference reviews. The search was limited to documents in English and was conducted on 23/05/2024 at 19:25. The whole analysis was conducted using Python, which provided the necessary tools and libraries for data handling, text processing, and advanced analytical techniques. A Text mining technique was employed to analyze the abstracts and extract relevant information, identifying and quantifying the frequency of occurrence of the keywords within each abstract, resulting in a database containing the accuracy of keyword occurrences across the corpus. Moreover, to analyze the impact of Industry 4.0 on the economic, social, and environmental dimensions of sustainability based on the factors that constitute each dimension, we conducted a bigram analysis. Sentiment analysis was then conducted to examine the pattern of the influence of Industry 4.0 on the environmental, economic, and social dimensions of sustainability by analyzing the polarity of the text, revealing the overall percentage of positive and negative impacts of Industry 4.0 on the three dimensions of sustainability. Finally, association rule mining was employed to investigate the probability of the arrangement and co-occurrence of the dimensions in relation to Industry 4.0, allowing for the identification of patterns and relationships between the different dimensions and Industry 4.0, providing insights into the interconnectedness and potential interdependencies among the variables under study. The combination of these three methods provided a comprehensive and robust approach to analyzing the influence of Industry 4.0 on the environmental, economic, and social dimensions of sustainability by leveraging a large corpus of academic content and employing advanced analytical techniques.

3.2. Data collection

The data collection approach for this study involved accessing Scopus.com and conducting a search using the keywords “industry 4.0,” “environmental,” “economic,” and “social dimensions of sustainability.” This search strategy ensured that the retrieved results were relevant to the core topic of examining the influence of Industry 4.0 on the environmental, economic, and social dimensions of sustainability.

A total of 6759 article abstracts were downloaded from Scopus.com based on the keyword search. The selection of Scopus.com as the data source is justified by its reputation as a comprehensive and reliable academic database, ensuring the credibility and quality of the retrieved abstracts.

The systematic keyword search approach demonstrates the methodological commitment and integrity to the outcomes. By carefully crafting the search query and selecting appropriate keywords, this strategy assisted in obtaining a representative sample of abstracts pertaining to the influence of Industry 4.0 on environmental, economic, and social dimensions of sustainability. The systematic structure of the data-collecting process improves the reliability and accuracy of the outputs, allowing for more informed analysis and interpretation in the subsequent stages of the study.

The use of a large corpus of 6759 abstracts from Scopus.com contributes to the comprehensiveness and representativeness of the data, enabling a thorough examination of the topic from various perspectives and sources. This rigorous data collection approach, which involved directly accessing and retrieving relevant abstracts from a reputable academic database, lays a solid foundation for the subsequent analysis through text mining, sentiment analysis, and association rule mining techniques.

3.3. Methods used

3.3.1. Text mining

For the text analysis approach, a method was developed to extract the keywords from the abstracts. The process begins by utilizing necessary tools for data manipulation and text processing, including tokenization for breaking text into individual words and stemming to reduce words to their root form. Subsequently, the method sets the predefined keywords $\mathrm{K} = \{\mathrm{k}_1, \mathrm{k}_2, \ldots \mathrm{k}_\mathrm{n}\}$ that will be used to search for and analyze their occurrences within a set of downloaded abstracts $\mathrm{A} = \{\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{m}\}$. The abstracts were previously retrieved from Scopus.com and stored in a CSV file.

The initial step involves calculating the total frequency count freq(k, A) of each keyword k across all abstracts in the set A [58]. This is achieved by summing up the individual frequency counts freq(k, a_i) of the keyword k in each abstract a_i, as represented by the equation:

$$freq(k,\ A)\ = \sum_{o} \leq i \leq m\ freq(k,\ a_i)$$

Here, freq(k, a_i) represents the frequency count of keyword k in the individual abstract a_i, as defined by the equation:

$$\boldsymbol{freq}(\boldsymbol{k},\ \boldsymbol{a})\ =\ \boldsymbol{\Sigma}\ \boldsymbol{I}(\boldsymbol{k}\ \in\ \boldsymbol{a})$$

where I(k ∈ a) is an indicator function that takes the value 1 if keyword k is present in abstract a, and 0 otherwise. By summing up these individual frequency counts across all abstracts, the algorithm obtains the total frequency count freq(k, A) for each keyword k in the set K [59].

After calculating the frequency counts for all keywords, the process proceeds to organize the results into a structured format. This can be mathematically represented as a matrix M of size $m \times n$, where each element M[i, j] represents the frequency count of keyword k_j in abstract a_i [60], as defined by the equation:

$$\boldsymbol{M}[\boldsymbol{i},\ \boldsymbol{j}]\ =\ \boldsymbol{freq}(\boldsymbol{k\_j},\ \boldsymbol{a\_i})$$

The data frame provides a tabular representation of the matrix M, with rows corresponding to abstracts and columns corresponding to keywords, facilitating further analysis or processing of the keyword frequency data.

3.3.2. Unsupervised categorization of abstracts into sustainability dimensions using keyword frequency analysis

To explore the impact of Industry 4.0 on economic, social, and environmental factors, we classified the abstracts into three dimensions of sustainability. This involved employing an automated method that examined the content of each abstract to identify whether its primary emphasis was on economic, social, or environmental considerations [61]. The process began with loading the abstracts into a data processing framework, denoted as *D*.

$$\boldsymbol{D} = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots\cdots, \boldsymbol{a}_n\}$$

Where a_i is an abstract in the dataset

Next, we assigned sustainability's economic, social, and environmental dimensions as keywords.

Ke="economic'
Ks="social"
Kenv="enviromental"

These keywords served as the basis for determining the primary focus of each abstract.

Each abstract a_i was then examined to identify its main focus based on the occurrence of the defined keywords. We converted the abstract text to a uniform format and counted the occurrences of each keyword within the text. $f(a_i, K)$ denote the frequency function that counts occurrences of the keyword $\boldsymbol{K}$ in the abstract a_i [62].The scores were computed as follows:

$$\boldsymbol{S}_e(\boldsymbol{a}_i) = \boldsymbol{f}(\boldsymbol{a}_i, \boldsymbol{K}_e)$$

$$\boldsymbol{S}_s(\boldsymbol{a}_i) = \boldsymbol{f}(\boldsymbol{a}_i, \boldsymbol{K}_s)$$

$$\boldsymbol{S}_{env}(\boldsymbol{a}_i) = \boldsymbol{f}(\boldsymbol{a}_i, \boldsymbol{K}_{env})$$

The keyword with the highest score determined the categorization of each abstract.

Abstracts containing more than one keyword with equal frequencies were excluded from the analysis to ensure clarity and focus on singular sustainability dimensions.

3.3.3. Bigram analysis

To analyze the impact of Industry 4.0 on economic, social, and environmental dimensions of sustainability based on the factors that constitute each dimension, we conducted a bigram analysis on separate datasets comprising scientific abstracts categorized into these dimensions.

The bigram analysis is grounded in text mining and NLP techniques, which involve extracting meaningful patterns and insights from large volumes of unstructured text data. This approach allows for the identification of frequent word pairs (bigrams) that co-occur within the text, providing insights into common themes and concepts. The text data from the abstracts were preprocessed to ensure consistency and accuracy. This involved tokenization, which breaks down the text into individual words, and the removal of common English stop words to focus on meaningful content. Stemming was also applied to reduce words to their root form, ensuring that variations of a word are treated as a single entity. Bigrams, which are pairs of consecutive words, were generated separately for each sustainability dimension (economic, social, and environmental). This step helps in capturing the context in which specific terms are used together, providing a deeper understanding of the relationships between concepts.

Mathematically, let $\boldsymbol{T}_{\boldsymbol{economic}}$, $\boldsymbol{T}_{\boldsymbol{social}}$, and $\boldsymbol{T}_{\boldsymbol{environmental}}$ represent the concatenated texts for economic, social, and environmental dimensions respectively:

$$\boldsymbol{T}_{economic} = \boldsymbol{concatenate}(\boldsymbol{Abstract}_i|\boldsymbol{S}_e)$$

$$\boldsymbol{T}_{social} = \boldsymbol{concatenate}(\boldsymbol{Abstract}_i|\boldsymbol{S}_s)$$

$$\boldsymbol{T}_{evvironmental} = \boldsymbol{concatenate}(\boldsymbol{Abstract}_i|\boldsymbol{S}_{env})$$

where $\boldsymbol{Abstract}_i$ are individual abstracts and each $\boldsymbol{S}$ denotes their assigned category.

Each text $\boldsymbol{T}_{economic}$, $\boldsymbol{T}_{social}$, and $\boldsymbol{T}_{environmental}$ was tokenized into individual words, and common English stop words were removed to focus on meaningful content. Let $\boldsymbol{W}_{economic}$, $\boldsymbol{W}_{social}$, and $\boldsymbol{W}_{environmental}$ represent the filtered lists of words after tokenization for each respective dimension.

Bigrams, which are pairs of consecutive words, were then generated separately from $\boldsymbol{W}_{economic}$, $\boldsymbol{W}_{social}$, and $\boldsymbol{W}_{environmental}$. Let $\boldsymbol{B}_{economic}$, $\boldsymbol{B}_{social}$, and $\boldsymbol{B}_{environmental}$, represent the sets of bigrams for economic, social, and environmental dimensions, respectively:

$$\boldsymbol{B}_{economic} = \{(\boldsymbol{w}_i, \boldsymbol{w}_{i+1}) | \boldsymbol{w}_i, \boldsymbol{w}_{i+1} \in \boldsymbol{W}_{economic}\}$$

$$\boldsymbol{B}_{social} = \{(\boldsymbol{w}_i, \boldsymbol{w}_{i+1}) | \boldsymbol{w}_i, \boldsymbol{w}_{i+1} \in \boldsymbol{W}_{social}\}$$

$$\boldsymbol{B}_{environmental} = \{(\boldsymbol{w}_i, \boldsymbol{w}_{i+1}) | \boldsymbol{w}_i, \boldsymbol{w}_{i+1} \in \boldsymbol{W}_{environmental}\}$$

The frequency of each bigram in $\boldsymbol{B}_{economic}$, $\boldsymbol{B}_{social}$, and $\boldsymbol{B}_{environmental}$ were calculated using Python's Counter class from the collection's module, and the 20 most common bigrams were identified for each dimension [62,63].

To assess the relevance of these bigrams to each sustainability dimension, we classified the top 20 bigrams based on their contextual meaning and their association with economic, social, or environmental factors. This classification facilitated an understanding of how Industry 4.0 impacts each dimension by analyzing bigram frequency and context.

By applying this methodological approach, we systematically processed the text data to extract and analyze frequent bigrams, providing nuanced insights into the distinct impacts of Industry 4.0 on economic, social, and environmental dimensions of sustainability. This approach is supported by similar text-mining techniques used in various studies to explore large corpora of academic literature and extract meaningful patterns [64,65].

3.3.4. Text sentiment analysis

For text sentiment analysis, a new algorithm was developed that uses the TextBlob library in Python, which employs the Pattern Analyzer algorithm. Based on the work of Kamps et al. [66], this algorithm measures the semantic orientation (positive or negative sentiment) of adjectives using WordNet, a lexical database of English words. The polarity score calculation in TextBlob follows this equation:

$$\textbf{polarity\_score} = \frac{\boldsymbol{P}(\textbf{positive}) - \boldsymbol{P}(\textbf{negative})}{\boldsymbol{P}(\textbf{positive}) + \boldsymbol{P}(\textbf{negative}) + 0.000001}$$

In this equation, P(positive)P(positive) and P(negative)P(negative) represent the probabilities of the text being classified as positive or negative, respectively. The small constant (0.000001) in the denominator prevents division by zero. These probabilities are derived from a naive Bayes classifier trained on a labeled dataset of positive and negative text samples.

The function uses the polarity score from TextBlob to assign a sentiment label ('positive', 'negative', or 'neutral') based on the following conditions:

- If the polarity score is greater than 0, the text is classified as having a positive sentiment.
- If the polarity score is <0, the text is classified as having a negative sentiment.
- If the polarity score is exactly 0, the text is classified as having a neutral sentiment.

TextBlob identifies sentiment polarity based on a predefined lexicon and a set of rules. Here's how it works:

For a positive sentiment example, consider the sentence: "The influence of Industry 4.0 on environmental sustainability is significant." Here, the words used have positive connotations, leading to a positive polarity score, which results in the classification of the text as having a positive sentiment. The sentiment calculation can be represented mathematically as follows:

$$\boldsymbol{P} = \frac{1}{\boldsymbol{n}} \sum_{i=1}^{n} \boldsymbol{s}(\boldsymbol{wi})$$

Where $\frac{1}{n}$ is a normalization factor representing the reciprocal of the total number of words in each sentence. This normalization ensures that the polarity score P is an average sentiment score per word, making it comparable across texts of different lengths. n is the number of words in the text, and (wi) is the sentiment score of the word wi.

For example, the influence of Industry 4.0 on sustainability has been transformative and has led to significant economic and social improvements. In this case, words such as "transformative" and "improvements" carry positive connotations.

Conversely, for a negative sentiment example, consider the sentence: "The impact of Industry 4.0 on environmental sustainability has been disastrous and has led to economic and social challenges." In this case, words such as "disastrous" and "challenges" carry negative connotations.

This approach allows the algorithm to classify the text as having a positive or negative sentiment toward the influence of Industry 4.0 on environmental, economic, and social dimensions of sustainability.

The polarity score in Text Blob is calculated based on a predefined lexicon and pattern analysis. A positive polarity score indicates positive sentiment, a negative score suggests negative sentiment and a score of zero indicates neutral sentiment. This simple rule-based system allows TextBlob to classify the sentiment of a given text automatically.

3.3.5. Association rule mining

Association rule mining was employed to reveal significant relationships as well as the probability and strength of these variables Industry 4.0, environmental, economic, and social dimensions of sustainability occurring together. The model first started by transforming the data into binomial, then utilizing the FP-Growth algorithm to uncover significant relationships among the variables: "industry 4.0," "environmental," "economic," and "social." The FP-Growth algorithm then identifies frequent item sets, ensuring each itemset meets a minimum support threshold denoted by (σ)

$$\boldsymbol{\sigma} = \frac{|\{\boldsymbol{t} \in \boldsymbol{T} | \boldsymbol{fi} \subseteq \boldsymbol{t}\}}{|\boldsymbol{T}|}$$

Where:

- σ: This is the support threshold. It represents the support value of a specific item set fi, indicating the proportion of transactions in the dataset that contain this item set. Support is a measure of how frequently an item set appears in the dataset.
- $\{t \in T | fi \subseteq t\}$: This is the set of transactions tt within the total transaction set T that contains the item set fi. The notation $fi \subseteq t$ means that all items in the itemset fi are present in the transaction tt.
- $|\{t \in T | f \subseteq t\}|$: This represents the cardinality (or size) of the set of transactions that contain the item set fi. It counts how many transactions are in the dataset including all items in fi.
- $|T|$: This is the total number of transactions in the dataset. It provides a reference for determining the proportion of transactions that include the itemset fi.

Subsequently, association rules are generated from these item sets, represented as $X \rightarrow$Y, where X and Y are disjoint subsets of items. The rules are assessed based on their support and confidence.

$$\textbf{Support } (\boldsymbol{X} \rightarrow \boldsymbol{Y}) = \frac{|\{\boldsymbol{t} \in \boldsymbol{T} | \boldsymbol{X} \cup \boldsymbol{Y} \subseteq \boldsymbol{t}\}}{|\boldsymbol{T}|}$$

$$\textbf{Confidence } (\boldsymbol{X} \rightarrow \boldsymbol{Y}) = \frac{\textbf{support}(\boldsymbol{X} \cup \boldsymbol{Y})}{\textbf{support}(\boldsymbol{X})}$$

Moreover, the probability of the variables coming together is assessed through the measure of lift, which indicates the strength of the association between variables. The lift of a rule X→Y is calculated as:

$$\textbf{Lift}\ (\boldsymbol{X}\rightarrow\boldsymbol{Y}) = \frac{\textbf{support}(\boldsymbol{X}\cup\boldsymbol{Y})}{\textbf{support}(\boldsymbol{X})\textbf{xsupport}(\textbf{Y})}$$

This measure helps determine whether the occurrence of *X* increases the likelihood of *Y* occurring together with it, relative to their occurrences. A lift value greater than 1 indicates a positive association, meaning that *X* and *Y* are more likely to occur together than would be expected if they were independent. The concept and calculation of lift are derived from the work by Agrawal et al. [67], which provides a comprehensive framework for association rule mining and related measures.

4. Results

The application of text mining, sentiment analysis, and association rule mining techniques to a large corpus of academic abstracts has yielded insightful results, shedding light on the intricate interplay between Industry 4.0 and the economic, social, and environmental dimensions of sustainability. Through these advanced analytical approaches, we have uncovered the dominant themes, prevailing sentiments, and intricate relationships that characterize this multifaceted domain.

Fig. 2 presents a word cloud visualization generated from the most frequently occurring meaningful words in the abstracts. The central theme, “industry 4.0,” is prominently featured, with surrounding keywords such as “economic,” “social sustainability,” “environmental,” “energy,” “efficiency,” “growth,” “development,” “supply chain,” and “new technologies.” The size and placement of these words highlight their significance and the interconnected nature of Industry 4.0′s impact across various domains. This visualization effectively encapsulates the key themes and concepts related to Industry 4.0, illustrating its multifaceted influence on economic, social, and environmental dimensions of sustainability.

Table 1, provides a categorization of the retrieved abstracts related to Industry 4.0 into three sustainability dimensions: Economic, Social, and Environmental. Additionally, there is a category for abstracts that did not fit into any of these dimensions or contain more than one dimension with equal frequencies and are thus labeled as Uncategorized. The categorization was performed based on the presence and frequency of specific keywords within each abstract.

The table categorizes 6752 scientific abstracts related to Industry 4.0 into three sustainability dimensions: Economic, Social, and Environmental, along with an Uncategorized category for abstracts that did not fit into any of these dimensions. The Economic category has the highest number of abstracts 2762, indicating a strong research focus on the economic implications of Industry 4.0, such as cost reduction and economic benefits of automation. The Social category includes 1728 abstracts, highlighting research on social impacts like employment and societal well-being. The Environmental category contains 1533 abstracts, reflecting studies on sustainable manufacturing and environmental conservation. There are 729 Uncategorized abstracts, which either span multiple dimensions equally or do not explicitly focus on any specified dimension, underscoring the interdisciplinary nature of Industry 4.0 research.

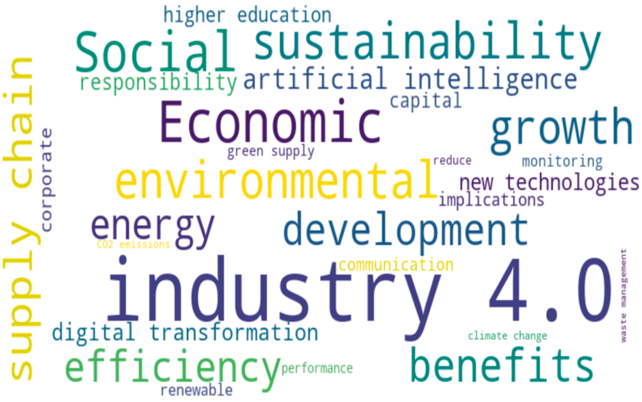


Fig. 2. A Word Cloud Analysis of Industry 4.0, Economic, Social, and Environmental Dimensions of Sustainability.
Source: Authors.

Table 1
Distribution of Abstracts by Sustainability Dimension.

Category	Count of Abstracts
Economic	2762
Social	1728
Environmental	1533
Uncategorized	729

Source: Authors.

Figs. 3–5 present treemap visualizations of bigram frequencies across the economic, social, and environmental dimensions of sustainability influenced by Industry 4.0. Fig. 3 highlights the economic dimension, with “Industry 4.0″ as the most frequent bigram (3717 occurrences). Significant factors include “Economic Growth” (718), “Economic Development” (666), and “Sustainable Development” (452), among others. This visualization emphasizes the various economic aspects impacted by Industry 4.0, such as “Energy Efficiency” (371) and “Energy Consumption” (267), illustrating a diverse range of economic implications. Fig. 4 focuses on the social dimension, with “Industry 4.0″ appearing 2945 times. Key topics include “Social Sustainability” (419), “Social Responsibility” (339), and “Education 4.0″ (309). Other notable bigrams are “Sustainable Development” (285) and “Social Networks” (247), reflecting themes like social dynamics and technological integration. This figure underscores the broad social impacts of Industry 4.0. Fig. 5 illustrates the environmental dimension, with “Industry 4.0″ as the dominant bigram (1984 occurrences). Important terms include “Environmental Sustainability” (341), “Environmental Impacts” (329), and “Energy Efficiency” (225). Topics like “Climate Change” (192) and “Waste Management” (168) highlight environmental conservation efforts. This visualization showcases the significant environmental research and practices influenced by Industry 4.0.

The stacked bar graph in Fig. 6, illustrates the sentiment analysis results related to Industry 4.0 and its impact on various sustainability dimensions. Each bar represents a different category industry 4.0, economic, social, and environmental, and is divided into sections indicating positive green and negative red sentiments. For Industry 4.0, the tallest bar shows a substantial volume of 6023 abstracts, with a predominant positive sentiment of 5564 abstracts and a smaller portion of negative sentiment of 459 abstracts. The economic dimension, with 2761 abstracts, displays a majority of positive sentiments 2608 abstracts, and fewer negative sentiments 153 abstracts. Similarly, the social dimension, comprising 1728 abstracts, reveals 1604 positive and 124 negative sentiments. The environmental dimension, represented by 1527 abstracts, shows a strong positive sentiment of 1352 abstracts compared to a lesser negative sentiment of 175 abstracts. The graph highlights the positive sentiment towards Industry 4.0 across all dimensions, with only a small fraction of abstracts expressing negative views.

Fig. 7 and Table 2 represent the results of an association rule mining analysis that aims to uncover frequent patterns and associations between Industry 4.0 and sustainability dimensions such as economic, social, and environmental aspects. The nodes in the figure show the tested concepts, while the edges connecting them represent association rules, indicating the relationships and co-occurrences between the concepts. Each edge is labeled with a rule number and two numerical values corresponding to support and confidence metrics. Support

Fig. 3. Tree map of Economic dimension Bigram Counts. Source: Authors.

measures the proportion of records containing both the antecedent and consequent concepts, while confidence indicates the proportion of records with the antecedent concept that also includes the consequent. Rule 3 (0.444/0.550) between “economic” and “industry 4.0″ suggests a 44.4 % support, meaning 44.4 % of records mention both concepts, and a 55.0 % confidence, implying that among records mentioning “economic,” 55.0 % also reference “industry 4.0.” Table 2 complements the figure by presenting the association rule metrics, listing the premises (antecedents), conclusions (consequents), support, confidence, and lift values. The lift metric measures the strength of the association rule, with a value greater than 1 indicating a positive correlation. The lift of Rule 3 which combines both “economic” and “industry 4.0″ shows a result of 1.060. Rule 5 (0.430/0.831) between “social” and “industry 4.0″ in the figure indicates 43.0 % support and 83.1 % confidence, implying a strong association between Industry 4.0 and social aspects. The row with the premise “Social” and conclusion “Industry 4.0″ in Table 2 presents the corresponding support of 0.430, confidence of 0.831, and a lift of 1.030 for this rule. Similarly, Rule 4 (0.380/0.827) between “environment” and “industry 4.0″ suggests 38.0 % support and 82.7 % confidence, highlighting the association between Industry 4.0 and environmental considerations. Table 2 shows a support of 0.380, a confidence of 0.827, and a lift of 1.024 for this rule. The combined analysis from Fig. 5 and Table 2 reveals strong associations between Industry 4.0 and economic, social, and environmental aspects,

Fig. 4. Tree map of Social Dimension Bigram Counts. Source: Authors.

indicating the multifaceted nature of discussions surrounding this topic. The quantitative metrics provide insights into these associations' strength and prevalence, underscoring the interdependence levels between Industry 4.0 and each sustainability dimension in academic discourse.

Further analysis of Table 2 reveals intricate relationships between Industry 4.0 and sustainability indicators through association rule metrics. The Support values identify the most frequent co-occurring patterns, while Confidence values indicate the strength and reliability of these relationships. By examining rules with high Support (≥0.5) and Confidence (≥0.7), we identified particularly strong associations that demonstrate robust integration between Industry 4.0 and sustainability aspects. The analysis revealed several important patterns: rules with high Confidence but moderate Support (0.3–0.5) suggest specialized but reliable relationships, while those with high Support but moderate Confidence (0.5–0.7) indicate commonly co-occurring elements that may have contextual dependencies. To validate these relationships, we employed additional metrics such as Lift (to measure correlation strength beyond random co-occurrence) and Conviction (to assess implication strength). This multi-metric approach provides a more nuanced understanding of the interconnections between Industry 4.0 technologies and sustainability indicators. Particularly strong rules

Fig. 5. Tree map of Environmental Dimension Bigram Counts. Source: Authors.

(Confidence >0.8) suggest potential causal relationships or strategic implementation patterns that could inform future sustainability initiatives in Industry 4.0 contexts. The varying levels of Support and Confidence across different rule sets also highlight the hierarchical nature of these relationships, where certain technological implementations appear to consistently precede or coincide with specific sustainability outcomes.

5. Discussion of key findings

This study investigates the interplay between Industry 4.0 and sustainability by applying advanced text mining, sentiment analysis, and association rule mining to analyze 6759 abstracts from the Scopus database. The findings reveal that Industry 4.0 is largely perceived positively across economic, social, and environmental dimensions, underscoring its potential as a transformative driver of sustainable development.

The novelty of this study lies in its holistic approach, as it simultaneously addresses the environmental, economic, and social dimensions of sustainability, providing a more integrated understanding of how Industry 4.0 influences sustainable development. Additionally, this study combines both qualitative and quantitative approaches. The qualitative aspect is addressed through text mining, bigram analysis, and sentiment analysis, while the quantitative aspect is achieved

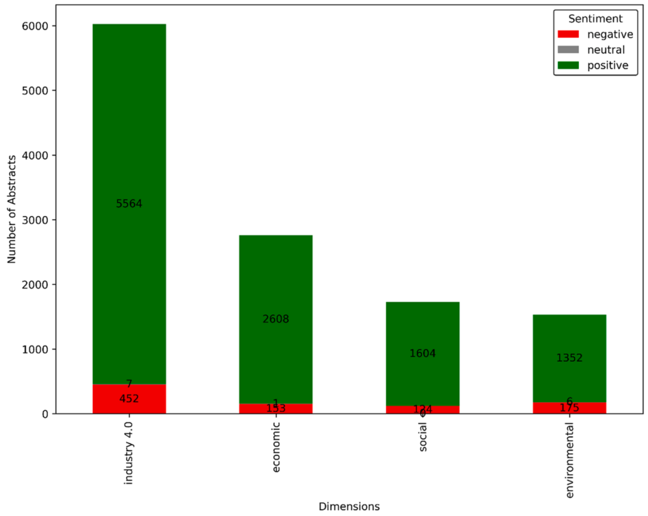


Fig. 6. Sentiment Distribution by Sustainability Dimensions. Source: Authors.

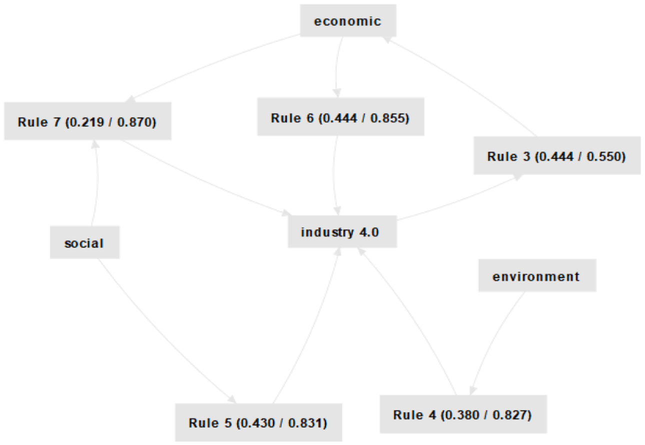


Fig. 7. Association rule mining of industry 4.0 with economic, social, and environmental aspects.
Source: Authors.

Table 2
Association rule metrics relating to industry 4.0 and its economic, social, and environmental impacts.

Premises	Conclusion	Support	Confidence	Lift
Economic	Industry 4.0	0.444	0.855	1.060
Social	Industry 4.0	0.430	0.831	1.030
Environment	Industry 4.0	0.380	0.827	1.024
Economic, Social	Industry 4.0	0.219	0.870	1.078

Source: Authors.

through association rule mining. This important transfer of qualitative data into quantitative data is facilitated by the co-occurrence of keywords.

In comparison to previous studies such as those by Beltrami et al. [1], Furstenau et al. [20], and Ejsmont et al. [21], this study provides a more detailed quantitative analysis alongside a robust empirical examination. While the prior works are primarily qualitative, focusing on reviews, conceptual frameworks, and bibliometric analyses, this study adopts a mixed-methods approach, integrating both qualitative and quantitative techniques.

The qualitative aspect of this study is addressed through text mining, bigram analysis, and sentiment analysis, which uncover patterns and perspectives on Industry 4.0′s impact on sustainability. The quantitative aspect is achieved through association rule mining, allowing for precise identification of the intensity and co-occurrence of sustainability themes with Industry 4.0 concepts. This transfer of qualitative data into quantitative data is made possible by the co-occurrence of keywords, creating a data-driven perspective on the relationships between Industry 4.0 and sustainability dimensions.

In contrast, studies like Beltrami et al. [1] utilized conceptualization and theorization to explore Industry 4.0′s sustainability potential, contributing valuable theoretical insights but without empirical testing. Similarly, Ejsmont et al. [21] conducted a bibliometric review of Industry 4.0′s impact, offering a broad research landscape overview but not investigating specific associations and sentiments as deeply as the present study's methods allow.

By combining text mining with sentiment and association analyses, this study offers a comprehensive, data-driven perspective on the multifaceted relationships between Industry 4.0 and sustainability, going beyond the limitations of previous qualitative-focused studies.

Table 3 summarizes the results of the bigram analysis, sentiment analysis, and association rule mining, shedding light on the relationships between Industry 4.0 and each sustainability dimension.

The analysis reveals a robust relationship between Industry 4.0 and economic sustainability, supported by frequent co-occurrence of terms such as “economic growth” and “circular economy” in the bigram analysis (Fig. 3), which appear 718 and 666 times, respectively. These high-frequency terms reinforce the perception that Industry 4.0 contributes significantly to economic prosperity while promoting sustainability. Sentiment analysis further highlights the positive outlook towards Industry 4.0′s economic impact, with 94.5 % (2608 out of 2761) of instances exhibiting positive sentiment. This suggests a shared understanding of Industry 4.0′s ability to drive economic growth, enhance resource efficiency, and facilitate waste reduction. The association rule mining results emphasize this relationship, with economic sustainability displaying a high support value of 0.444, confidence of 0.855, and lift of 1.060, indicating a strong and reliable association. These findings align with the role of Industry 4.0 in fostering new business models, optimizing supply chains, and enabling circular economy practices, thereby advancing both economic outcomes and sustainability goals. This is consistent with the conclusions drawn by Herrmann et al. [10] and Kiel et al. [11], who emphasized the economic benefits of Industry 4.0 in relation to resource optimization.

The social dimension of Industry 4.0 presents a more comprehensive view, capturing both its opportunities and risks for societal impacts. The bigram analysis (Fig. 4) reveals frequent associations with terms like “social responsibility” and “education 4.0,” suggesting that Industry 4.0′s social sustainability impact extends beyond worker safety and job creation to include broader aspects like inclusivity and connectivity. The sentiment analysis shows a positive perspective in 92.8 % (1604 out of 1728) of the cases, highlighting optimism about Industry 4.0′s potential to improve workplace safety, create new high-tech jobs, and address social equity. The association rule mining findings validate this view, with a support value of 0.430, confidence of 0.831, and lift of 1.030, indicating a moderately strong but positive association between Industry 4.0 and social sustainability. These results support the idea that Industry 4.0 can address societal needs, such as automating hazardous tasks, providing upskilling opportunities, and fostering socially responsible practices, aligning with the insights of Beier et al. [16] and Stock et al. [13] on the mixed social impacts of automation.

In the environmental dimension, the findings suggest a meaningful association between Industry 4.0 and environmental sustainability, as evidenced by frequent co-occurrence of terms such as “energy efficiency” and “waste management” (Fig. 5). The sentiment analysis reveals 88.5 % (1352 out of 1527) positive sentiment, reflecting a favorable view of Industry 4.0′s role in optimizing energy use, reducing emissions, and minimizing waste. The association rule mining results show a support value of 0.380, confidence of 0.827, and lift of 1.024,

Table 3
Summary of analytical findings on the interplay between industry 4.0 and sustainability dimensions.

Dimension	Bigram analysis	Sentiment analysis	Association rule mining		
			Support	Confidence	Lift
Industry4.0-Economic	“economic growth,” “circular economy”	2608 positive out of 2761	0.444	0.855	1.060
Industry4.0-Social	“social responsibility,” “education 4.0″	1604 positive out of 1728	0.430	0.831	1.030
Industry4.0-Environmental	“energy efficiency,” “waste management	1352 positive out of 1527	0.380	0.827	1.024
Industry4.0-Economic & Social	“economic growth,” “social responsibility”	-	0.219	0.870	1.078

Source: Authors.

confirming a strong link to environmental sustainability, although slightly lower than the economic and social dimensions. These findings highlight the potential of Industry 4.0 technologies, such as IoT and big data analytics, to advance eco-friendly practices by enhancing energy efficiency and enabling real-time environmental monitoring. This aligns with prior research by Bai et al. [24] and Ghobakhloo [23], which emphasized the positive environmental impact of Industry 4.0 in reducing carbon emissions and waste.

Expanding on this multifaceted impact, the analysis highlights a synergistic relationship between Industry 4.0 and the combined economic and social dimensions of sustainability. The co-occurrence of terms such as “economic growth,” “social responsibility,” “education 4.0,” and “workplace safety” reveals a holistic approach where Industry 4.0 is perceived as a driver of both economic prosperity and societal well-being. This relationship is validated by association rule mining results, which show a support value of 0.219, confidence of 0.870, and lift of 1.078 for the rule connecting Industry 4.0 with both economic and social dimensions. This high confidence level indicates a strong likelihood that discussions on economic sustainability often address social aspects as well, reflecting a common perspective of Industry 4.0 as a tool for achieving economic gains while enhancing social conditions. The lift value greater than 1.0 signifies a positive correlation, suggesting that Industry 4.0 discussions are frequently aligned with themes that promote economic inclusivity and social equity.

These combined findings highlight Industry 4.0′s potential to support sustainable development through initiatives like circular economy practices, worker upskilling, job creation in high-tech fields, and safer workplaces. As such, Industry 4.0 is not only seen as a means to drive efficiency and economic growth but also as a catalyst for social responsibility, fulfilling interconnected goals of economic stability and social well-being. This integrated perspective reinforces the view that Industry 4.0 is instrumental in creating an inclusive, sustainable future where technological advancement supports broader economic and social objectives.

6. Conclusion, implications and limitations

6.1. *Main findings*

Based on the comprehensive analysis of 6759 abstracts from the Scopus database, Industry 4.0 emerges as a transformative force with significant positive implications across economic, social, and environmental sustainability dimensions. The study’s novelty lies in its holistic approach, combining qualitative methods (text mining, bigram analysis, and sentiment analysis) with quantitative techniques (association rule mining) to provide an integrated understanding of Industry 4.0′s influence on sustainable development. This mixed-methods approach provides a more robust analysis compared to previous studies that primarily focused on qualitative reviews and bibliometric analyses.

The findings reveal that Industry 4.0 is largely perceived positively across all sustainability dimensions, with particularly strong associations in the economic sphere. The frequent co-occurrence of terms such as “economic growth” and “circular economy” emphasizes Industry 4.0′s potential to drive economic prosperity while promoting sustainability. The economic dimension shows the most robust relationship, suggesting that Industry 4.0 is primarily viewed as a catalyst for growth, efficiency, and new business opportunities. However, the significant attention to social and environmental dimensions demonstrates a growing recognition of the need for a holistic approach to sustainability in the context of technological advancement.

The social dimension of Industry 4.0 presents a clearer picture, capturing both opportunities and potential challenges. The analysis reveals frequent associations with terms like “social responsibility” and “Education 4.0,” indicating that Industry 4.0′s social impact extends beyond traditional considerations of workplace safety and job creation to encompass broader aspects of inclusivity and connectivity. The environmental dimension shows meaningful associations with concepts such as energy efficiency and waste management, highlighting Industry 4.0′s potential to advance eco-friendly practices through IoT and big data analytics.

6.2. *Practical implications*

The findings have significant implications for policymakers. Policies should be developed that address all three dimensions of sustainability simultaneously, recognizing the interconnected nature of economic, social, and environmental impacts. While supporting the economic potential of Industry 4.0, policy frameworks should incorporate strong protections for social and environmental sustainability. The emphasis on “Education 4.0″ suggests that particular attention should be paid to adapting educational systems and workforce training programs to prepare for the technological changes brought by Industry 4.0. Environmental regulations should leverage the potential of Industry 4.0 technologies to address challenges such as improving energy efficiency and waste management.

The interconnected nature of Industry 4.0 with various sustainability concepts emphasizes the complex and multifaceted nature of this technological revolution. This suggests that the implementation of Industry 4.0 technologies cannot be viewed in isolation but must be considered within the broader context of sustainable development. Regular assessments of social impacts should be conducted to ensure that technological advancements contribute positively to social sustainability. Given the positive view, increased funding for research into the sustainable implementation of Industry 4.0 technologies across all dimensions should be considered, along with policies that promote collaboration between industry, academia, and civil society.

6.3. *Limitations and future directions*

The study acknowledges several methodological constraints that present opportunities for future research. The utilization of Scopus database abstracts, while extensive, may not encompass the complete scope of Industry 4.0 and sustainability research. Further investigations would benefit from incorporating full-text articles across multiple databases to enhance analytical comprehensiveness.

Additionally, the current implementation of sentiment analysis and association rule mining, though systematic, operates within the parameters of predefined algorithms. This methodology may present limitations in capturing the full complexity of human sentiment and intricate relationship patterns. Future research could advance the analytical framework through the integration of sophisticated machine learning techniques and enhanced natural language processing capabilities.

Furthermore, while this research establishes a foundational understanding of Industry 4.0′s relationship with sustainability, it necessarily maintains a broad perspective rather than examining industry-specific dynamics. Subsequent research would benefit from conducting in-depth case studies and sector-specific analyses, thereby providing targeted insights for industry practitioners and policy development. Such focused investigations could yield more precise, actionable recommendations for specific industrial contexts.

Funding

Open Access funding provided by Hungarian Electronic Information Services National Programme (EISZ) – University of Dunaujvaros.

Ethics approval and consent to participate

The authors undertake that this article has not been published in any other journal and that no plagiarism has occurred.

CRediT authorship contribution statement

Mohamad Ali Saleh Saleh: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mutaz AlShafeey:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors have no relevant financial or non-financial interests to disclose.

Acknowledgments

We acknowledge that we have used AI tools to enhance the readability and clarity of the language used in this manuscript.

Data availability

Data will be made available on request.

References

[1] M. Beltrami, G. Orzes, J. Sarkis, M. Sartor, Industry 4.0 and sustainability: towards conceptualization and theory, J. Clean. Prod. 312 (2021) 127733.
[2] E. Hofmann, M. Rüsch, Industry 4.0 and the current status as well as future prospects on logistics, Comput. Ind. 89 (2017) 23–34.
[3] G. Culot, G. Nassimbeni, G. Orzes, M. Sartor, Behind the definition of Industry 4.0: analysis and open questions, Int. J. Product. Econ. 226 (2020) 107617.
[4] B.C. Group, Nine Technologies Transforming Industrial Production, Boston, 2019.
[5] W. Liao, T. Wang, A novel collaborative optimization model for job shop production–delivery considering time window and carbon emission, Sustainability 11 (10) (2019) 2781.
[6] F. Tao, H. Zhang, A. Liu, A.Y. Nee, Digital twin in industry: state-of-the-art, IEEE Transact. Ind. Inform. 15 (4) (2018) 2405–2415.
[7] I. Castelo-Branco, F. Cruz-Jesus, T. Oliveira, Assessing Industry 4.0 readiness in manufacturing: evidence for the European Union, Comput. Ind. 107 (2019) 22–32.
[8] UN World Commission on Environment and Development, Report of the World Commission On Environment and Development: Our Common Future, New York, 1987.
[9] J. Elkington, Partnerships from cannibals with forks: the triple bottom line of 21st-century business, Environ Qual. Manag. 8 (1) (1998) 37–51.
[10] C. Herrmann, et al., Sustainability in manufacturing and factories of the future, Int. J. Precis. Eng. Manufact.-Green Technol. 1 (2014) 283–292.
[11] D. Kiel, J.M. Müller, C. Arnold, K.-I. Voigt, Sustainable industrial value creation: benefits and challenges of industry 4.0, Int. J. Innov. Manag. 21 (08) (2017) 1740015.
[12] G. Bressanelli, F. Adrodegari, M. Perona, N. Saccani, Exploring how usage-focused business models enable circular economy through digital technologies, Sustainability 10 (3) (2018) 639.
[13] T. Stock, M. Obenaus, S. Kunz, H. Kohl, Industry 4.0 as enabler for a sustainable development: a qualitative assessment of its ecological and social potential, Process Saf. Environ. Protect. 118 (2018) 254–267.
[14] A. Bremer, Diffusion of the "Internet of Things" on the world of skilled work and resulting consequences for the man–machine interaction, Empir. Res. Vocat. Educ. Train. 7 (2015) 1–13.
[15] P. Isaias, T. Issa, V. Chang, T. Issa, Outlining the issues of cloud computing and sustainability opportunities and risks in European organizations: a SEM study, J. Electron. Commer. Organ. (JECO) 13 (4) (2015) 1–25.
[16] G. Beier, S. Niehoff, T. Ziems, B. Xue, Sustainability aspects of a digitalized industry–A comparative study from China and Germany, Int. J. Precis. Eng. Manufact.-Green Technol. 4 (2017) 227–234.
[17] H. Rong, et al., Optimizing energy consumption for data centers, Renew. Sustain. Energy Rev. 58 (2016) 674–691.
[18] W. Tushar, et al., Peer-to-peer energy systems for connected communities: a review of recent advances and emerging challenges, Appl. Energy 282 (2021) 116131.
[19] S. Ford, M. Despeisse, Additive manufacturing and sustainability: an exploratory study of the advantages and challenges, J. Clean. Prod. 137 (2016) 1573–1587.
[20] L.B. Furstenau, et al., Link between sustainability and industry 4.0: trends, challenges and new perspectives, IEEE Access 8 (2020) 140079–140096.
[21] K. Ejsmont, B. Gladysz, A. Kluczek, Impact of industry 4.0 on sustainability—Bibliometric literature review, Sustainability 12 (14) (2020) 5650.
[22] B. Purvis, Y. Mao, D. Robinson, Three pillars of sustainability: in search of conceptual origins, Sustain. Sci. 14 (2019) 681–695.
[23] M. Ghobakhloo, Industry 4.0, digitization, and opportunities for sustainability, J. Clean. Prod. 252 (2020) 119869.
[24] C. Bai, P. Dallasega, G. Orzes, J. Sarkis, Industry 4.0 technologies assessment: a sustainability perspective, Int. J. Product. Econ. 229 (2020) 107776.
[25] Henning, K., W. Wolfgang, and H. Johannes, *Recommendations for implementing the strategic initiative INDUSTRIE 4.0*. Final report of the Industrie, 2013. **4**(0).
[26] J.M. Tien, The next industrial revolution: integrated services and goods, J. Syst. Sci. Syst. Eng. 21 (2012) 257–296.
[27] A. Radziwon, A. Bilberg, M. Bogers, E.S. Madsen, The smart factory: exploring adaptive and flexible manufacturing solutions, Procedia Eng. 69 (2014) 1184–1190.
[28] H. Lasi, et al., Industry 4.0, Bus. Inform. Syst. Eng. 6 (2014) 239–242.
[29] E. Commission, Factories of the Future: multi-Annual Roadmap for the Contractual PPP under Horizon 2020, 2013.
[30] F. Mosconi, The New European industrial policy: Global competitiveness and the Manufacturing Renaissance, Routledge, 2015.
[31] P.L. Hartman, J.A. Ogden, J.R. Wirthlin, B.T. Hazen, Nearshoring, reshoring, and insourcing: moving beyond the total cost of ownership conversation, Bus. Horiz. 60 (3) (2017) 363–373.
[32] M. Ghombakhloo, The future of manufacturing industry: a strategic roadmap towards Industry 4.0, J. Manuf. Technol. Manag 29 (6) (2018) 910–936.
[33] G.L. Tortorella, D. Fettermann, Implementation of Industry 4.0 and lean production in Brazilian manufacturing companies, Int. J. Product. Res. 56 (8) (2018) 2975–2987.
[34] J.D. Sachs, From millennium development goals to sustainable development goals, Lancet 379 (9832) (2012) 2206–2211.
[35] A. Colantonio, Social Sustainability: a Review and Critique of Traditional Versus Emerging Themes and Assessment Methods, 2009.
[36] B.S. Kumar, V. Ravi, A survey of the applications of text mining in financial domain, Knowl. Based Syst. 114 (2016) 128–147.
[37] V. Gupta, G.S. Lehal, A survey of text mining techniques and applications, J. Emerg. Technolog. Web Intell. 1 (1) (2009) 60–76.
[38] M. Tan, Q. Le, Efficientnet: rethinking model scaling for convolutional neural networks, in: International conference on machine learning, 2019, p. PMLR.
[39] H. Hassani, et al., Text mining in big data analytics, Big Data Cognit. Comput. 4 (1) (2020) 1.
[40] C.D. Manning, H. Schütze, Foundations of Statistical Natural Language Processing, MIT Press, Cambridge, MA, 2012.
[41] B. Liu, Sentiment Analysis and Opinion Mining, Springer Nature, 2022.
[42] B. Pang, L. Lee, Opinion mining and sentiment analysis, Found. Trend.® Inform. Retriev. 2 (1–2) (2008) 1–135.
[43] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: a survey, Ain Shams Eng. J. 5 (4) (2014) 1093–1113.
[44] E. Cambria, B. Schuller, Y. Xia, C. Havasi, New avenues in opinion mining and sentiment analysis, IEEE Intell. Syst. 28 (2) (2013) 15–21.
[45] J. Lee, H.-A. Kao, S. Yang, Service innovation and smart analytics for industry 4.0 and big data environment, Proced. Cirp. 16 (2014) 3–8.
[46] S.S. Kamble, A. Gunasekaran, S.A. Gawankar, Sustainable Industry 4.0 framework: a systematic literature review identifying the current trends and future perspectives, Process Saf. Environ. Protect. 117 (2018) 408–425.
[47] A.G. Frank, L.S. Dalenogare, N.F. Ayala, Industry 4.0 technologies: implementation patterns in manufacturing companies, Int. J. Product. Econ. 210 (2019) 15–26.
[48] D.L.M. Nascimento, et al., Exploring Industry 4.0 technologies to enable circular economy practices in a manufacturing context: a business model proposal, J. Manufact. Technol. Manag. 30 (3) (2019) 607–627.

[49] S. Saberi, M. Kouhizadeh, J. Sarkis, L. Shen, Blockchain technology and its relationships to sustainable supply chain management, Int. J. Product. Res. 57 (7) (2019) 2117–2135.

[50] S. Bag, S. Gupta, A. Kumar, U. Sivarajah, An integrated artificial intelligence framework for knowledge creation and B2B marketing rational decision making for improving firm performance, Ind. Market. Manag. 92 (2021) 178–189.

[51] C.B. Frey, M.A. Osborne, The future of employment: how susceptible are jobs to computerisation? Technol. Forecast. Soc. Change 114 (2017) 254–280.

[52] A. Benešová, J. Tupa, Requirements for education and qualification of people in Industry 4.0, Procedia Manufact. 11 (2017) 2195–2202.

[53] P. Fantini, M. Pinzone, M. Taisch, Placing the operator at the centre of Industry 4.0 design: modelling and assessing human activities within cyber-physical systems, Comput. Ind. Eng. 139 (2020) 105058.

[54] E. Rauch, C. Linder, P. Dallasega, Anthropocentric perspective of production before and within Industry 4.0, Comput. Ind. Eng. 139 (2020) 105644.

[55] P. Rosa, et al., Assessing relations between Circular Economy and Industry 4.0: a systematic literature review, Int. J. Product. Res. 58 (6) (2020) 1662–1687.

[56] A.B.L. de Sousa Jabbour, C.J.C. Jabbour, C. Foropon, M.G. Filho, When titans meet–Can industry 4.0 revolutionise the environmentally-sustainable manufacturing wave? The role of critical success factors, Technol. Forecast. Soc. Change 132 (2018) 18–25.

[57] Y. Cui, S. Kara, K.C. Chan, Manufacturing big data ecosystem: a systematic literature review, Robot. Comput. Integr. Manuf. 62 (2020) 101861.

[58] G. Casella, R.L. Berger, Statistical Inference, 2nd ed., Duxbury Press, Pacific Grove, 2002.

[59] W. McKinney, Python For Data Analysis: Data Wrangling With Pandas, NumPy, and IPython, 1st ed., O'Reilly Media, Inc, Sebastopol, CA, 2013.

[60] G. Strang, Introduction to Linear Algebra, 5th ed., Wellesley-Cambridge Press, Wellesley, MA, 2016.

[61] G. Andrienko, et al., Visual analytics of mobility and transportation: state of the art and further research directions, IEEE Transact. Intell. Transport. Syst. 18 (8) (2017) 2232–2249.

[62] D. Jurafsky, J.H. Martin, Speech and Language Processing, 3rd ed., Pearson, New York, NY, 2019.

[63] C.D. Manning, P. Raghavan, H. Schütze. *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, 2008.

[64] R. Feldman, J. Sanger, The Text Mining handbook: Advanced Approaches in Analyzing Unstructured Data, Cambridge university press, 2006.

[65] C.C. Aggarwal, C. Zhai, Mining Text Data, Springer Science & Business Media, New York, NY, 2012.

[66] J. Kamps, M. Marx, R.J. Mokken, M. De Rijke, Using WordNet to measure semantic orientations of adjectives. Lrec, Citeseer, 2004.

[67] R. Agrawal, T. Imieliński, A. Swami, Mining association rules between sets of items in large databases, in: Proceedings of the 1993 ACM SIGMOD international conference on Management of data, 1993.