https://doi.org/10.1186/s40537-023-00848-8

Journal of Big Data

RESEARCH Open Access

Big Data Analytics-based life cycle sustainability assessment for sustainable manufacturing enterprises evaluation

Lahcen Tamym[1*], Lyes Benyoucef[1], Ahmed Nait Sidi Moh[2] and Moulay Driss El Ouadghiri[3]

*Correspondence: lahcen.tamym@lis-lab.fr

[1] CNRS, LIS, Aix-Marseille University, University of Toulon, Marseille, France
[2] LASPI, IUT of Roanne, Jean Monnet University, Saint Etienne, Roanne, France
[3] IA Laboratory, Moulay Ismail University, Meknes, Morocco

Abstract

Recently, governments and organizations have repeatedly pressed manufacturing enterprises to promote the ethical and transparent use of natural resources, lessen their negative effects on national and international ecosystems, and safeguard people and the environment. In this context, enhancing the various stages of the product/service life cycle to fulfill sustainability requirements and foster sustainable value creation is a key area of interest for researchers and professionals. This emphasis reflects the growing recognition of the importance of minimizing the environmental impact of products and services, while also maximizing their positive contributions to society, economy, and environment. To this end, this research work addresses how manufacturing enterprises benefit from life cycle sustainability assessment (LCSA) thinking to incorporate the environmental and social criteria into the product/service life cycle strategies. To do so, a novel approach based on environmental priority strategy (EPS) as an LCSA method for impacts monetization coupling with Big Data Analytics (BDA) techniques and tools is developed to evaluate and analyze the manufacturing enterprises' impacts on the environment and society. Moreover, the developed approach evaluates manufacturing enterprises' progress toward sustainable development goals (SDGs). Finally, to demonstrate the applicability of the developed approach, a case study from the corporate environmental impact database is used, and the obtained numerical results are analyzed showing its efficiency and added value.

Keywords: Manufacturing enterprises, Sustainability, Big Data Analytics, Life cycle sustainability assessment, Environmental priority strategy, Sustainable development goals

Introduction

In the present era, manufacturing enterprises face ongoing pressure from governments and organizations to incorporate responsible and transparent practices into their business processes. This includes promoting the sustainable and efficient utilization of natural resources, minimizing adverse effects on local and global ecosystems, and ensuring the well-being and protection of individuals and communities. Therefore, manufacturing enterprises have to integrate sustainability thinking into their daily operations and activities. Hence, sustainability has gained popularity on a global scale from both enterprises and governments. Thus, it is necessary for governments and enterprises' stakeholders (such as customers, business partners, regulators, support organizations, etc.) to work together in order to achieve sustainable development [66]. Furthermore, manufacturing enterprises encounter numerous sustainability concerns that require their attention. These concerns encompass a wide range of issues, including climate change mitigation and adaptation, safeguarding human health and safety, addressing resource depletion and scarcity, managing fluctuating material and energy prices, complying with environmental protection legislation, and responding to social pressures for ethical and responsible business practices. These concerns highlight the complex and interconnected nature of sustainability challenges that manufacturing enterprises must navigate in order to achieve long-term viability and meet stakeholder expectations [18].

It is evident that integrating environmental and social considerations into the business models and decision-making frameworks of manufacturing enterprises presents notable challenges [12]. In this context, sustainable value creation (SVC) is considered the main challenge for manufacturing enterprises in the current business [58]. Hence, SVC thinking enables manufacturing enterprises to benefit from environmental and social resources while developing their economy without harming the planet and people. In this regard, manufacturing enterprises have to shift their SVC by prolonging the products' lifetime and gaining a long-term profit [71]. Additionally, this shift will alter production and consumption practices to be more environmentally friendly and decrease the overall amount of non-sustainable materials used by manufacturing enterprises [45]. *Generally, this research work focuses on the challenges of manufacturing enterprises in their SVC and the sustainability criteria integration throughout the product/service life cycles. This leads us to address the first research question: How do manufacturing enterprises integrate sustainability thinking into their production processes?.*

Improving the stages of the product/service life cycle to meet sustainability criteria is a problem that attracted many researchers and practitioners in the sustainable manufacturing enterprises field. This problem is mainly known as "life cycle sustainability assessment (LCSA)" [17, 24, 35]. The LCSA places a strong emphasis on evaluating the sustainability criteria, particularly, impacts, emissions, and natural resources utilization, which threaten the basic requirements of both present-day and future generations. Furthermore, it assesses various adverse impacts on the environment, society, and economy, while also enhancing the decision-making process for developing sustainable products/services, which is central to SVC. Moreover, numerous studies have explored the potential of Big Data Analytics (BDA) in promoting sustainability within manufacturing enterprises [58]. Examples of such applications can be found in the literature, including studies, such as value creation with BDA for enterprises [42], big data-driven framework for sustainable and smart additive manufacturing [38], and sustainable industrial value creation benefits and challenges of Industry 4.0 [31], among others. These studies highlight the diverse ways in which BDA is utilized to drive sustainability and enhance value creation in manufacturing enterprises contexts. *Nonetheless, as far as our knowledge extends, there is currently a lack of prior research investigating the combination of BDA potentials and LCSA methods for evaluating the sustainability of manufacturing enterprises. This represents notable research gaps and highlights the potential for future studies to explore the synergies and benefits that can be derived from integrating BDA and LCSA* *in assessing and improving the sustainability performance of manufacturing operations*, such as:

- Many studies focus on assessing environmental data for sustainability improvements. For instance, incorporating aspects such as worker well-being, community impacts, and social equity into the analysis could result in holistic sustainability strategies. Consequently, there is a gap in integrating social and environmental data to provide a full view of sustainability.
- Social impacts are always difficult to quantify and analyze. Hence, there is little understanding of how to incorporate life-cycle evaluation of product impacts during manufacturing enterprise processes using BDA models and approaches.
- Monitoring and decision-making in real-time using BDA could be extremely beneficial and enrich the literature by creating algorithms and frameworks that enable manufacturing enterprises to make quick sustainability-related decisions based on real-time data.
- Integrating supply chain management in the product life cycle of manufacturing enterprises is challenging. As a result, there's a gap in exploring how BDA can be used to enhance sustainability across the entire supply chain, from raw material sourcing to end-of-life product management.
- BDA applications for sustainability assessments are still in their infancy. Thus, it is unclear how to successfully achieve sustainability needs. Literally, achieving the 2030 Sustainable Development Goals (SDG) may be impossible for stakeholders in the key fields of sustainable development without a solid understanding of the above challenges.

From the above challenges and gaps, the present research work discusses the problem of manufacturing enterprises in maintaining the linkages between the environmental, and social pillars of sustainability. Moreover, impacts monetization is one of the best solutions to measure the manifold impacts on the environment and society. This end leads us to address the second research question: How can manufacturing enterprises couple environmental priority strategy (EPS) as an LCSA method with BDA techniques and tools to evaluate and analyze their impacts on both environment and society?

The rest of the paper is organized as follows: "Literature review" section starts by giving some definitions of sustainable manufacturing enterprises, and then presents how the use of BDA techniques and tools can promote their sustainable development. In addition, it briefly debates the use of LCSA methods in assessing the stage of product/service life cycle during manufacturing enterprises' activities and operations. "Theoretical backgrounds" section gives some definitions of EPS method and impact monetization. Moreover, it provides an overview of the BDA tools and techniques used in this study. While "BDA-based LCSA approach" section presents the proposed approach which combines BDA and LCSA methods. Besides, "Experiments results and analyses" section demonstrates the applicability of the developed approach, where a case study from the corporate environmental impact database is used, and the numerical results are analyzed. Finally, "Conclusion" section concludes the paper with some challenges and upcoming perspectives.

Literature review

This section reviews some research works related to the considered problem. Firstly, some definitions of sustainable manufacturing enterprises and the debate about the integration of sustainability into manufacturing enterprises' processes and operations are given. Secondly, the discussions on how BD and BDA can promote sustainable development of manufacturing enterprises are argued. Finally, the last part of the literature review consists of the definition of LCSA and impacts monetization, as well as, an overview of the incorporation of life cycle thinking into manufacturing enterprises is discussed.

Sustainable manufacturing enterprises

A sustainable manufacturing enterprise has been defined as the development and designing of sustainable products that are not economically-based profits only, but also the integration of environmental and societal performances into every stage of products' life cycles [23, 25]. Besides, minimizing the negative environmental and social impacts, conserving energy and natural resources, and enhancing employees' health and community safety are mandatory in recent businesses for sustainable manufacturing processes. Moreover, this allows them to incorporate environmental and social sustainability thinking into their businesses. Furthermore, the incorporation of sustainability metrics into the manufacturing enterprises' business models is indispensable [61]. This involves considering sustainability at every stage, from product design and development to production, distribution, consumption, and disposal [29]. Collaboration and coordination across different business fields are essential to foster sustainable practices and drive positive environmental and social impacts while ensuring long-term economic viability [61]. In this regard, Jamwal et al. [29] presented a systematic literature review study based on well-known scientific databases to discover and evaluate the evolution of contributions in Industry 4.0 technologies in achieving manufacturing enterprises' sustainability. Consequently, the authors concluded that: "*most of the conducted research in this field focused on general theatricals and conceptual, and just a few of them have studied the real applications of different technologies for achieving manufacturing enterprises' sustainability*". In fact, a majority of these studies often overlook the integration of social and environmental issues, which are fundamental aspects of the Sustainable Development Goals (SDGs) agenda set forth by the United Nations. The SDGs provide a comprehensive framework for addressing global challenges and achieving sustainable development by addressing economic, social, and environmental dimensions in an integrated manner [66]. Moreover, there are many issues to consider when evaluating the manufacturing enterprises' environmental and social sustainability performances, including energy consumption, water consumption, use of raw materials and resource depletion, workplace danger, and employees' health [34, 43]. For instance, An et al. [4] argued that manufacturing enterprises are facing significant pressures from governments, clients, and competitors to optimize their energy consumption. Besides, the authors considered that manufacturing systems' operation and maintenance are probably the majority concerned with energy optimization. *Thus, optimizing the energy among these activities will significantly improve the enterprise's sustainability.* Furthermore, similar studies are trying to evaluate the environmental and societal effects of greenhouse gas emissions and energy consumption related to the manufacturing systems during production processes (machining and non-machining activities), such as Afrin et al. [2] and Battaïa et al. [8].

Current business witnesses widespread global competition in terms of sustainable product design (SPD) to meet the increasing demand for customized and stable production, as well as, the integration of environmental and societal sustainability criteria into product design. Indeed, eco-design and flexibility are critical keys to developing modern products by integrating their whole life cycle from cradle to grave [63]. For instance, Zhou et al. [75] emphasized that sustainable product design issues are mainly related to the increasing consumer demands for sustainability performance. With the aim of supporting Sustainable Product Development (SPD) under epistemic uncertainty, the authors put forth a proposed approach utilizing a multi-objective optimization-based technique known as the Order Preference by Similarity to the Ideal Solution (TOPSIS). The rationale behind employing this technique lies in its effectiveness in handling the epistemic uncertainty associated with customer preferences concerning optimization objectives. The authors sought to leverage TOPSIS as an efficient tool to address the challenges posed by uncertain customer preferences in the optimization process within SPD. Moreover, in the circular economy model, product designers and engineers have to consider the environmental and social performance of the product's life cycle. Thus, these kinds of products should be flexible in their design to be reused and recycled [14]. This can be done by evaluating the impact that initial design decisions have on each stage of the product's life cycle, such as product design, raw material sourcing, manufacturing, matching and non-machining operations and services, and end-of-life stages.

Digital technologies, including the Internet of Things (IoT), have significantly transformed the way businesses and industries operate, including their impact on environmental and social analytics through automating life cycle inventory (LCI) processes [19]. LCI is a critical component of assessing the environmental and social impacts of products and services across their entire life cycle, from raw material extraction to end-of-life disposal. IoT devices play a crucial role in automating data collection throughout a product's life cycle. These devices can be embedded in various stages of production, distribution, and usage to gather real-time data on factors such as energy consumption, emissions, water usage, and product location. This automated data collection enhances the accuracy and granularity of LCI, enabling more comprehensive environmental and social impact assessments. In addition, IoT-generated data can be analyzed in real-time, allowing for immediate identification of anomalies or deviations from sustainable practices [53]. This timely information empowers businesses to take corrective actions promptly, reducing negative environmental and social effects. In conclusion, digital technologies like IoT have a transformative role in automating life cycle inventory for environmental and social analytics. By enhancing data collection, process optimization, transparency, and consumer engagement, IoT contributes to more comprehensive and accurate assessments of a product's environmental and social impacts, ultimately driving more sustainable business practices [47]. *Consequently, implementing greener production technology for manufacturing enterprises' systems for developing new tools and methodologies is necessary so that engineers, designers, and manufacturers can benefit* *from it in order to make socially and eco-friendly responsible decisions throughout the product's life cycle.*

Big Data Analytics and sustainable development

Nowadays, the digital world knows an exponential increase in data volume around many domains. For instance, United Nations, [67] claimed that the size of the created data was 64.2 zettabytes in 2020 which is a 314% increase from 2015. Accordingly, they also argued that the COVID-19 pandemic contributed largely to creating this amount of data due to the amount of technology used during the pandemic period. Furthermore, this overwhelming amount of data is generated through the routine interactions with digital products and services, such as mobile phones, credit cards, and social media. These everyday activities contribute to the continuous flow of data that holds valuable insights and potential for analysis and interpretation. Thus, big data (BD) is distinguished by the five great Vs (i.e., volume, veracity, variety, velocity, and value) [20, 21]. BDA has emerged as the new and predominant tool for processing and analyzing these massive volumes of data, aiming to extract valuable insights and ensure the credibility and reliability of Big Data sources. BDA techniques enable organizations to effectively handle the complexity and scale of Big Data, facilitating the extraction of meaningful information and actionable intelligence from diverse data sources . By applying advanced analytical algorithms and methodologies, BDA empowers decision-makers to derive valuable insights, make informed decisions, and gain a deeper understanding of the data-driven landscape [59]. *Furthermore, BDA enables sustainable enterprises to improve their data-driven decision-making systems to better adapt to their environment and gain competitive advantages in the global market* [46, 60]. BDA supports enterprises' business by providing new and efficient methods that can deal with massive volumes of data to boost enterprises' efficiency and transparency through their operations and activities. Undoubtedly, BDA's role in a business process is to settle the limitations of traditional analysis of data by permitting enterprises more agility to understand situations and address related issues, as well as, to process heterogeneous data in real-time [62]. Moreover, the advances in computing capabilities of BDA techniques and platforms provide high-quality information that is detailed, timely, and relevant such as data mining (DM), machine learning (ML), and business intelligence (BI).

In 2015 and 2022, the United Nations launched two programs for sustainability development. The first program named "the 2030 Agenda for sustainable development" includes planning actions for people, the planet, and the prosperity of nations [66]. Indeed, this agenda envisages goals and targets that are ambitious and transformational visions for a sustainable world by considering economical, environmental, and social challenges in order to achieve the SDGs [66, 74]. Also, these SDGs are known as 17 grand challenges (GC) [72]. While the second program called "BD for sustainable development" intends to benefit from the potential and opportunities that BD has brought to society, especially the business development side [67]. Furthermore, this program demonstrates how organizations, enterprises, and governments can harness the power of appropriate BD to promote global, regional, and national sustainable development and work towards achieving the SDGs by 2030. By effectively utilizing BD analytics and insights, stakeholders can make informed decisions, develop targeted strategies, and implement impactful initiatives that contribute to sustainable development objectives. The program highlights the potential of BD as a valuable resource in driving positive change and aligning efforts towards the SDGs agenda within various geographic scales and contexts [10]. *For instance, United Nations claimed that the policy-makers, decision-makers, and business leaders can exploit and analyze BD of the poorest and marginalized populations using BDA techniques and tools in order to end extreme poverty and reach zero emissions by 2030. Obviously, in the current business analytics, sustainability and BDA are two key challenges to a successful business in highly competitive markets.* Thus, sustainable enterprises are aware of integrating sustainability thinking and BDA technology into their business models and supply chain strategies [36]. Moreover, embracing this new thinking will empower organizations to develop environmentally-friendly products and deliver sustainable services that not only meet customer expectations but also outperform their competitors. By integrating sustainability principles into their business models and operations, organizations can enhance their competitiveness by catering to the growing demand for green and sustainable offerings. This proactive approach not only aligns with customer preferences but also contributes to a positive brand image, fosters customer loyalty, and creates a distinct competitive advantage in the market. Accordingly, sustainability analytics is referred to as the use of BDA to achieve sustainable development and the incorporation of social and environmental responsibility into a business process [39]. Furthermore, according to Martí and Anna, [40], BDA technology will promote economical, environmental, social, ethical, legal, and political sustainability benefits for enterprises. *As a result, enterprises can benefit from this huge potential of BDA technology in terms of collecting, processing, and analyzing massive BD in real-time (or near real-time) in order to evaluate the sustainability factors. This can be done by generating the right insights and information to support decision-making throughout their operations and activities. Undoubtedly, this will enable manufacturing enterprises to better achieve sustainability-related initiatives and improve overall resource efficiencies, such as energy and resource use, greenhouse gas emissions, human health issues, and logistics performances, to name a few. Finally, this scenario analysis will keep enterprises maintaining the linkages between the economical, social, and environmental pillars of sustainability.*

Life cycle sustainability assessment and impacts monetization

The industrial sector is known for its extreme effects on the society and environment, due to the significant quantities of emissions that are being released, the waste of large amounts of materials, and the remarkable depletion of natural resources. These effects have resulted from their daily activities and operations. In fact, enterprises in the current business are aware of integrating sustainability metrics into their sustainable business models and sustainable supply chains [61]. In doing so, they have to improve the areas of products/services' life cycle to meet sustainability needs. *For this reason, the concept of "Life cycle sustainability assessment (LCSA)" refers to assessing all sorts of negative impacts on the environment, society, and economy, as well as, improving the decision-making process for new sustainable products/services throughout their life cycles* [24]. LCSA is a standard model that integrates the triple-bottom-line or the three pillars of sustainability (i.e., economic, environmental, social) into the life cycle of products/

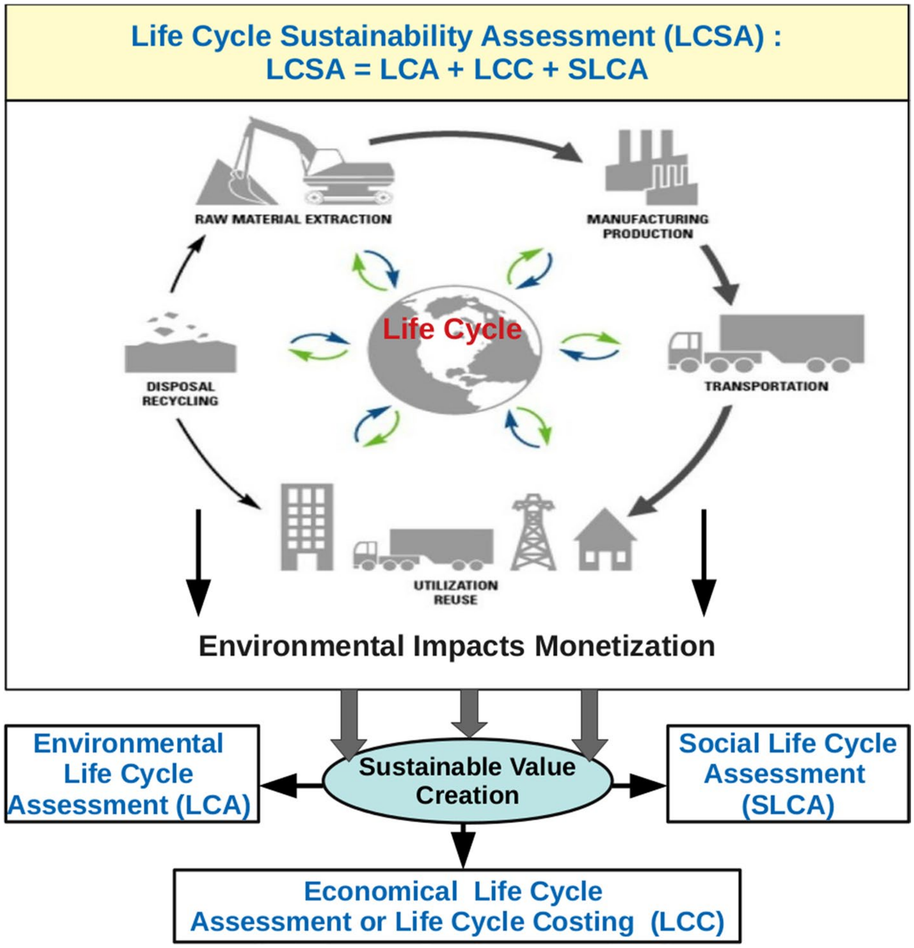

Fig. 1 Life cycle sustainability assessment

services. Moreover, this integration covers all life cycle stages from raw materials to the end user. In addition, LCSA keeps these three pillars aligned with each other, as shown in Fig. 1. Where, environmental life cycle assessment (LCA) is the environmental LCSA dedicated to assessing the environmental impacts that are resulted from commercial products, processes, or services [35]. For instance, for manufacturing enterprises, the environmental impacts are evaluated from raw material extraction and processing (cradle), manufacturing process, distribution and consumption, recycling and reuse, or final disposal of components (grave) [58]. Then, life cycle costing (LCC) is the economic side of LCSA, also known as, full-cost accounting or total-cost assessment [33]. While social life cycle assessment (SLCA) incorporates the social aspect into traditional LCSA that focuses only on environmental and economic aspects [48]. *Hence, the LCSA model enables manufacturing enterprises to increase sustainable practices within every stage of the product/service life cycle. Accordingly, it makes their value creation (VC) models cover the full impacts on the indirect stakeholders (i.e. planet and people)* [61].

Standardizing sustainability measures is highly difficult. In fact, it's difficult to determine which environmental and social impacts are the most severe, to be prioritized and reduced [5]. Monetization of impacts is one of the best solutions to measure the manifold impacts on the environment and society. In a broader sense, it can be used in the creation of sustainable value by enterprises for non-marketed goods, such as ecosystem services and biodiversity. Thus, *the LCSA methods focus on the quantification and determination of the associated economic value of impacts that are resulted from the emissions released by harmful substances, as well as, the exhaustion of non-sustainable resources during the life-cycle of a product or service. LCSA also known as life cycle analysis or eco-balance is a methodology dedicated to assessing the impacts related to every product stage in its life cycle (i.e., from raw materials, manufacturing, distribution, consumption, maintenance, repairing, recycling* [57, 65]. Therefore, several LCSA monetization methods were developed in the literature, namely Ecovalue09 [3], Stepwise2006 [13], LIME3 [70], Ecotax [16], EVR [68], and EPS [56].

From the above literature review, even if several efforts have been achieved to develop methodologies dealing with manufacturing enterprises' sustainability issues, there is a real need to come up with new approaches that integrate the three sustainability pillars using new technologies, such as BDA. Moreover, recognizing the potential of BDA technology, the monetization impacts of the EPS method with BDA techniques and tools are coupled in this study. Thus, this coupling enables managers to analyze and evaluate manufacturing enterprises' sustainability. More specifically, the present work focuses on the use of the EPS method for monetary values calculation of environmental and social impacts. This is done on the five general human basic needs, such as human health, ecosystem production, abiotic resources, and biodiversity. Besides, these basic needs are the most concern in the sustainability evaluation for the environment and society. In addition, the impact on these basic needs contributes significantly to the inhibition of sustainable development progress.

Theoretical backgrounds

Environmental priority strategy

Definitions

Environmental priority strategy (EPS) was developed in 1989 in a joint effort between the Volvo Car Company, the Sweden Institute for Environmental Research, and the Sweden Federation of Industries, and since then it has been revised several times in the context of projects [64]. The idea behind developing such a method is the great demand for sustainable products. Thus, EPS makes it easier for product developers to make the best decisions in the process of designing their products. It provides them with the comparison criteria with old reference products and gives indications of what is good versus bad. *Moreover, EPS allows measurement and evaluation of the chain of cause-effect and the cost of environmental and social damage* [26]. *These measurements and evaluation are according to the endpoints and the impacts monetization based upon the ISO standard and the LCA concept* (see Fig. 2). Since then, the method was changed several times, and the most recent version was released in 2015 [64].

Impact assessment in EPS

According to Steen, [55], the EPS method is based on five critical basic needs to human well-being called "safeguard subjects", namely, human health, ecosystem production capacity, abiotic resource, bio-diversity, and cultural and recreational values. Moreover,

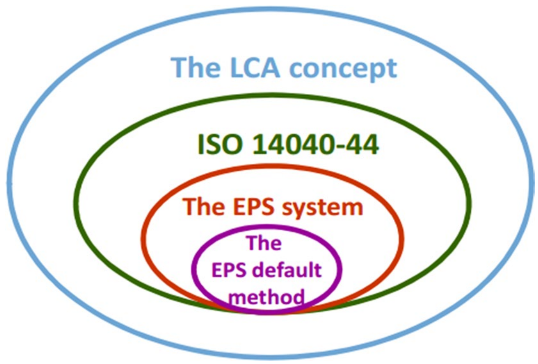

Fig. 2 EPS methodology as describe by its founder [55]

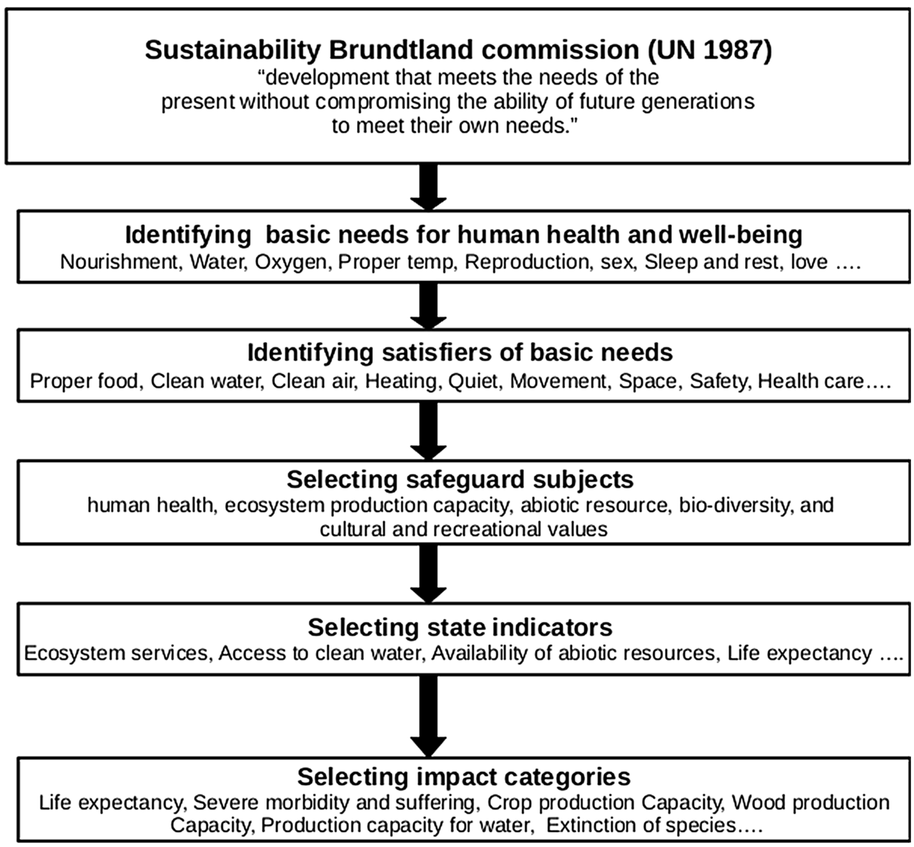

Fig. 3 Top-down approach for selecting safeguard subjects, state indicators, and impact categories

these safeguard subjects are the major economical, environmental, and social areas concerned by protection in order to meet SDGs launched by the United Nations. Thus, describing the impacts on the satisfiers of these critical basic needs to human well-being, a selection of the impact categories and category indicators were performed in the EPS method as illustrated in Appendix 1. Fig. 3 depicts this selection following a top-down approach. Furthermore, it begins by defining the sustainability concepts and pointing out the critical human basic needs to be met. In addition, the identification of relevant satisfiers for each basic need (safeguard subjects) is performed. Finally, a selection of state indicators series (or indicators of the endpoint category) which are appropriate for the impacts description of the product's life cycle on these basic needs is made.

Each safeguard subject is subject to a wide range of impacts that comes from daily activities in society. And indicators measure different impacts flows called pathways or endpoints effects, as described in Appendix 2. These pathways have resulted from effects related to emissions or resource depletion, for instance, CO_2 emissions to air affect global warming and climate change which in their turn will affect several safeguard subjects. Besides, evaluating environmental and social impacts is essential to estimate the damage that daily human activities have on people and planets. Hence, one way to assign monetary values to the effects brought on by ecologically damaging substances or excessive exploitation of natural resources is the impacts monetization [5]. Thus, market values are used to estimate the monetary values of the impacts on safeguard subjects that are critical to human basic needs. While the non-renewable stock of natural resources can be replaced by sustainable alternatives in a sustainable society. The cost of this replacement can be estimated based on the individuals' "Willingness to pay (WTP)" to produce a sustainable alternative or to avoid environmental damage. In addition, more details about the monetary values of environmental impacts on the five safeguard subjects are given in Steen [56, 57].

Big Data Analytics models and techniques

Nowadays, enterprises generate a great variety of data along their operations and activities. This critical velocity of data generation creates challenges for enterprises in terms of storage, processing, and computing [59]. Thus, specific and powerful data storage platforms were developed to support a new kind of database architectures and computing capabilities, such as Apache Spark, Hadoop, NoSQL (i.e. Cassandra and MongoDB), Enterprise Data Warehouses, Big Data Warehouse, Talend for Big Data, and Cloud computing and storage (i.e. Amazon Web Services, Google Cloud, Microsoft Azure). In addition, BD technologies show their efficiency and potential to manage BD that can come from a variety of sources in structured, semi-structured, and unstructured formats [7, 20], such as textual data, images, videos, online blogs, IoT sensors data, climate data, environmental data, social media data, consumer reviews, to name a few. Moreover, extracting accurate and trustworthy data is a big challenge for enterprises in order to derive insights and support decision-making processes. To this end, since data is available as the raw material, according to Arunachalam et al. [7] enterprises should exploit the high-tech facilities of BDA to build powerful analytics systems. Otherwise, they will fall into "Data Poor and Information Poor". Thus, making this possible, after generating the right raw data there are critical steps to follow as described below.

BD integration and management (DIM)

During their businesses, enterprises are generating a great amount of data from heterogeneous and various sources. In this regard, DIM is enterprises' ability to integrate, transform, and store the different types of heterogeneous raw data collected across their activities and processes in near/real-time into a single database. Moreover, this process of integration can be defined as the acquisition, pre-processing, and storage of data, respectively [11]. Thus, DIM is the most important step in the BD value chain because, it solves the inconsistencies and conflicts in terms of semantic, structure, and syntactic issues during the data fusion process for consistent business intelligence or analytics [76]. For instance, the integration and loading of the data in a BD warehouse for online analytical processing OLAP systems is performed using tools, such as Extract Transform Load (ETL)/Extract Load Transform (ELT). Besides, in-memory storage and processing of data in distributed and parallel manners are superior to conventional data storage and processing, since, they offer robust and quick querying of data as well as large-scale processing and storage of various data kinds in batch-wise, or near/real-time [27].

BD advanced analytics and visualization

After the data get integrated and prepared in a single and unified database, advanced analytics skills are needed to derive valuable insights and information for decision-making. To this end, enterprises need to leverage the analytical tools and techniques to analyze and process their BD either using batch size or real-time/near-time splitting. According to Souza, [54], the analytics is performed through four critical levels, namely descriptive, diagnostic, predictive, and prescriptive analytics. Therefore, descriptive analytics provides information on what happened based on historical data. Descriptive analytics can be usually performed using classical statistics methods and data mining (DM) techniques to see what happened. At the same time, diagnostic analytics enables one to evaluate what happened and give meaningful information by studying the causes of trends and correlations between variables. While predictive analytics uses machine learning (ML) algorithms and models to predict future events and what could happen based on historical, or real/near-time data [44]. Furthermore, this level of analytics can be achieved through the combined utilization of statistics, data mining (DM), and machine learning (ML). Prescriptive analytics, on the other hand, plays a vital role in selecting the optimal and most favorable option from a range of solutions, providing guidance for future outcomes. This level of analytics aids decision-makers in planning and optimizing both operational and strategic decisions pertaining to daily activities and functions of an enterprise. The ultimate goal of BDA is to disrupt traditional analysis by enabling agility in understanding and addressing issues through real-time processing of diverse and simultaneous data sources. For example, ML and deep learning technologies can be leveraged to accomplish this objective [62]. It is important to note that while Business Intelligence (BI) techniques, including data visualization and sophisticated analytical approaches like DM, may be employed, they often fall short in handling unstructured data. Therefore, specific procedures and strategies exclusive to BD must be employed [52, 59].

To the best of our knowledge, Apache Spark has been dedicated to being in front of BDA by including unified frameworks for batch and streaming processing and integration [37]. *Consequently, in this research study, Apache Spark is used to implement scalable solutions for BD preparation, integration routines, advanced analytics, and visualization.*

BDA-based LCSA approach

This section describes in detail the different components of the proposed approach. As shown in Fig. 4, it contains three main components, namely, environmental and social impacts monetization, Apache Spark ecosystem, and BDA for LCSA.

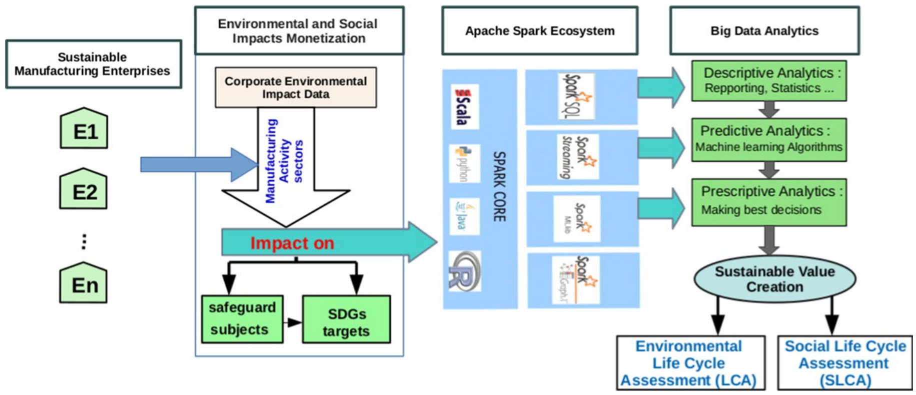

Fig. 4 Framework of the proposed BDA-based life cycle sustainability assessment approach

Table 1 Used sample data

Database name	# of manufacturing activity sectors	# of manufacturing enterprises	# of countries	# of observations
Corporate environmental impacts	20	839	41	5100

Environmental and social impacts monetization

To demonstrate the efficiency of BDA in the evaluation of manufacturing enterprises' sustainability, a real database is used (to access the database, the reader can connect by using the link [22]). Thus, the inputs of the proposed approach are data from the Impact Weighted Accounts Project (IWAP) developed by a team at Harvard Business School [50]. They adopted a methodology that makes use of a number of reputable academic sources and publicly available data to determine monetized estimates and measures of the organizational environmental and social impacts from processes and activities around the globe. The main goal behind this methodology is to show business leaders how to calculate the environmental impacts and rectify day-to-day decision-making. *Furthermore, this will allow manufacturing enterprises to manage risks and use natural resources sustainably*. These measurements pertain to the impacts on "safeguard subjects" [9] which include factors such as human health (working capacity) (WC), crop production capacity (CropPC), meat production capacity (MeatPC), fish production capacity (FishPC), wood production capacity (WoodPC), drinking water and irrigation water (water production capacity) (WaterPCDI), abiotic resources, and biodiversity. Additionally, the impacts of emissions were calculated for the 17 relevant SDGs targets by linking each emission's characterization pathways to a specific SDG target. The definitions of each SDG are provided in Appendix 3. For a more comprehensive understanding of how these environmental impact measurements are derived from organizational operations, readers are referred to the methodology developed by the IWAP team [50]. These measurements include data that differs from conventional environmental evaluations commonly used by investors and stakeholders, and this information holds significant value.

In this case study on sustainable manufacturing enterprises evaluation, Table 1 *illustrates the used sample data. Within this sample, 839 manufacturing enterprises operating* *in 20 manufacturing activity sectors located in 41 different countries were selected. It's important to note that only cases where the discount rate is set at 0% were considered.*

Adapted Apache Spark ecosystem

In order to execute, maintain, and create reliable pipelines and algorithms for large-scale data analysis across a diversity of workloads, BDA requires efficient frameworks. Fortunately, in order to solve BD challenges using a single processing tool and multi-functional and flexible languages, Apache Spark for BDA has come to light as a unifying engine for data science and engineering [49]. Additionally, Apache Spark is renowned for its distributed computing and cutting-edge in-memory processing programming tool. Additionally, it is regarded as the most active BD open-source project and is a quick and scalable framework [51]. In contrast to disk-based models like Hadoop's MapReduce, Apache Spark is an in-memory multistage programming tool. Moreover, it integrates a number of languages to provide extensive and effective APIs (i.e. Python, Java, Scala, SQL, and R). This integration and merging between APIs enable Apache Spark to perform complex distributed computing and storage. The Apache Spark engine is designed to exceed batch applications by including powerful computations that require separate distributed systems, such as iterative algorithms, interactive queries, and streaming [32]. Consequently, Apache Spark is more suitable for BD analysis than Apache Hadoop because its system is composed of two levels, namely, Spark core and Upper-Level Libraries, where, each level includes several components.

Spark core

Using a straightforward programming interface and a cluster of nodes, spark core enables the processing of massive BD sets. This interface offers effective fundamentals for data sharing between computations and is known as the Resilient Distributed Datasets (RDDs) concept. Indeed, a read-only, and partitioned collection of records is what an RDD is, according to Zaharia, [73]. Hence, users can directly store data on disk or in memory, manage its partitioning, and manipulate it using a wide range of operators thanks to RDDs' fault-tolerant, parallel data structures. *Thus, to support functions like data transformations and in-memory cluster computing, Java, Python, and R APIs were integrated into the development of the spark core, which was created using the Scala programming language.* Additionally, the spark core makes it possible to perform other functions like dynamic memory, task scheduling, data shuffling, and fault remediation [28].

Upper-level libraries

Upper-level libraries within Apache Spark include Spark's MLlib, GraphX, Spark Streaming, and Spark SQL. Spark's MLlib enables the implementation of large-scale ML algorithms based on data or model parallelism using Spark core [41]. Moreover, building ML algorithms require implementing machine-usual tasks like extracting and transforming features, training the model, evaluating the model, etc. For doing so, to create such effective algorithms and pipelines, Spark's MLlib is developed as a distributed ML library. Regarding this, Spark's MLlib is made up of two main packages, namely, spark.mllib and spark.ml [49]. Whereas, spark.mllib is constructed on top of RDDs to provide APIs for setting up, debugging, and tuning ML pipelines. While spark.ml, which has packages for linear algebra, statistics, and other fundamental ML functions, is developed on top of DataFrames. Besides, GraphX is a library for scalable graph analysis on Apache Spark [69]. It offers many functionalities for representing graph-oriented data based on graph transformations, graph algorithms, and graph builders [49]. Thus, these powerful functionalities provided by GraphX give Apache Spark a unified framework for all operators, algorithms, and pipelines that are involved in graph data representations and graph-distributed computation steps. While Apache Spark's stream processing is made to handle large-scale and real-time analysis. specifically, it permits Apache Spark to act on BD as soon as it arrives [30]. Finlay, Spark SQL offers a programming framework called DataFrame that enables Apache Spark to process structured data and can function as a distributed and scalable SQL query engine. [6].

Experiments results and analyses

This section discusses the third component of the proposed BDA-based life cycle sustainability assessment approach. More specifically, it presents the experiments and numerical results obtained using the Apache Spark ecosystem on corporate environmental impact data analysis. In this regard, three levels of BDA analytics are conducted on the selected sample of manufacturing enterprises, as shown in Fig. 4. Thus, this section demonstrates how BDA can be applied to evaluate manufacturing enterprises LCSA which are the main goals of this study.

Descriptive and diagnostic analytics

In the corporate environmental impact database, the calculated environmental impacts were performed from 2010 to 2019. Figure 5 shows the number of reported monetization impacts of manufacturing enterprises that are considered for environmental impact calculations. The increasing observations over considered years indicate that manufacturing enterprises are aware of environmental damage and the depletion of critical resources in their business and sustainable development. Thus, manufacturing enterprises are making publicly disclosing their environmental and social impacts data.

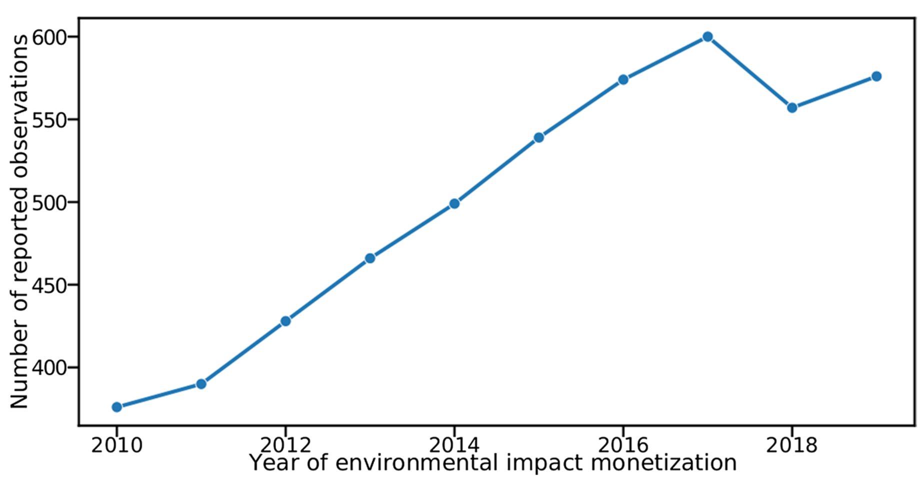

Fig. 5 Evolution of environmental impact monetization calculated for manufacturing enterprises

In the selected manufacturing activity sectors, each sector contains a significant number of manufacturing enterprises of different sizes. Each activity sector absolutely has different impacts on the environment and society. Thus, in this part of the analysis, the calculation of the total environmental cost (TEC) of each sector is performed based on the sum of the TEC of manufacturing enterprises in that sector. Figure 6 delineates the TEC of each manufacturing activity sector from 2010 to 2019. Consequently, sectors, such as "manufacture of medical, precision and optical instruments, watches and clocks", "manufacture of wearing apparel; dressing and dyeing of fur", "chemical industry", and "production of electricity" have high TEC in the selected sample. Furthermore, even if these four sectors have few manufacturing enterprises, they exercise significant environmental damage. This can be explained by the nature of the resources and procedures that they are using in their manufacturing systems, as well as, the poor consideration of sustainability criteria evaluation in their products' life cycles.

As described in the EPS method (see Fig. 3), there are five safeguard subjects: human health, ecosystem production capacity, abiotic resource, bio-diversity, and cultural and recreational values [55]. In the IWAP database, four of these safeguard subjects were considered by the monetization of the impacts exercised by manufacturing enterprises from their products' life cycles. Appendix 1 indicates each safeguard subject and its group of impact categories. However, only one impact category for each of the safeguard subjects is considered. Figures 7 and 8 show the impacts evolution of each manufacturing activity sector on each safeguard subject over the years. *It is important to note that the selected sample was limited to human health working capacity, water production capacity (irrigation water & drinking water), abiotic resources, and biodiversity. Other impact categories, such as wood production capacity, meat production capacity, and crop*

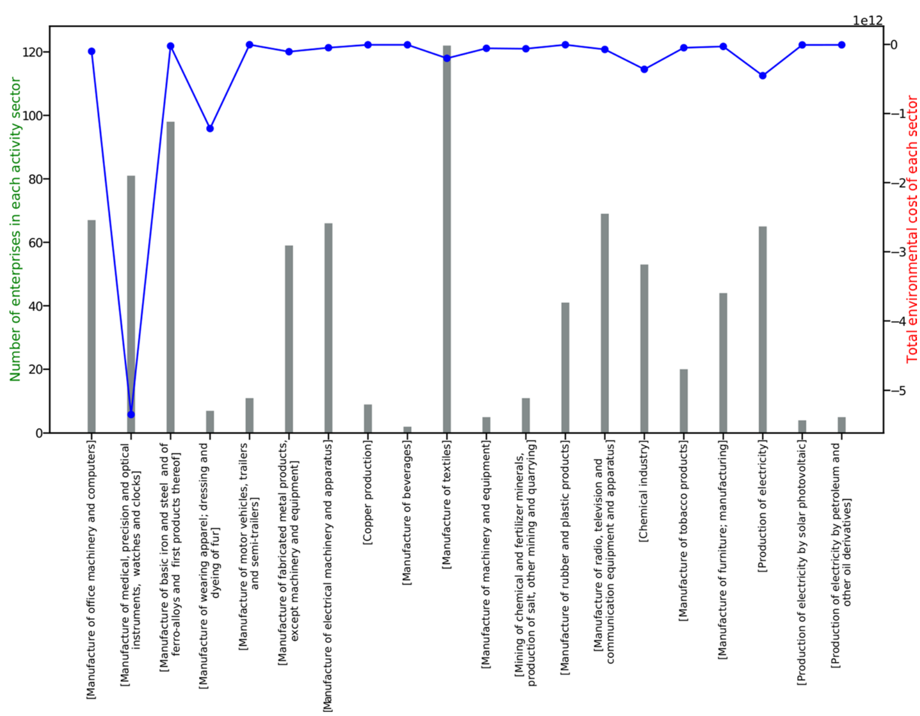

Fig. 6 Total environmental cost incurred by manufacturing enterprises in each manufacturing activity sector

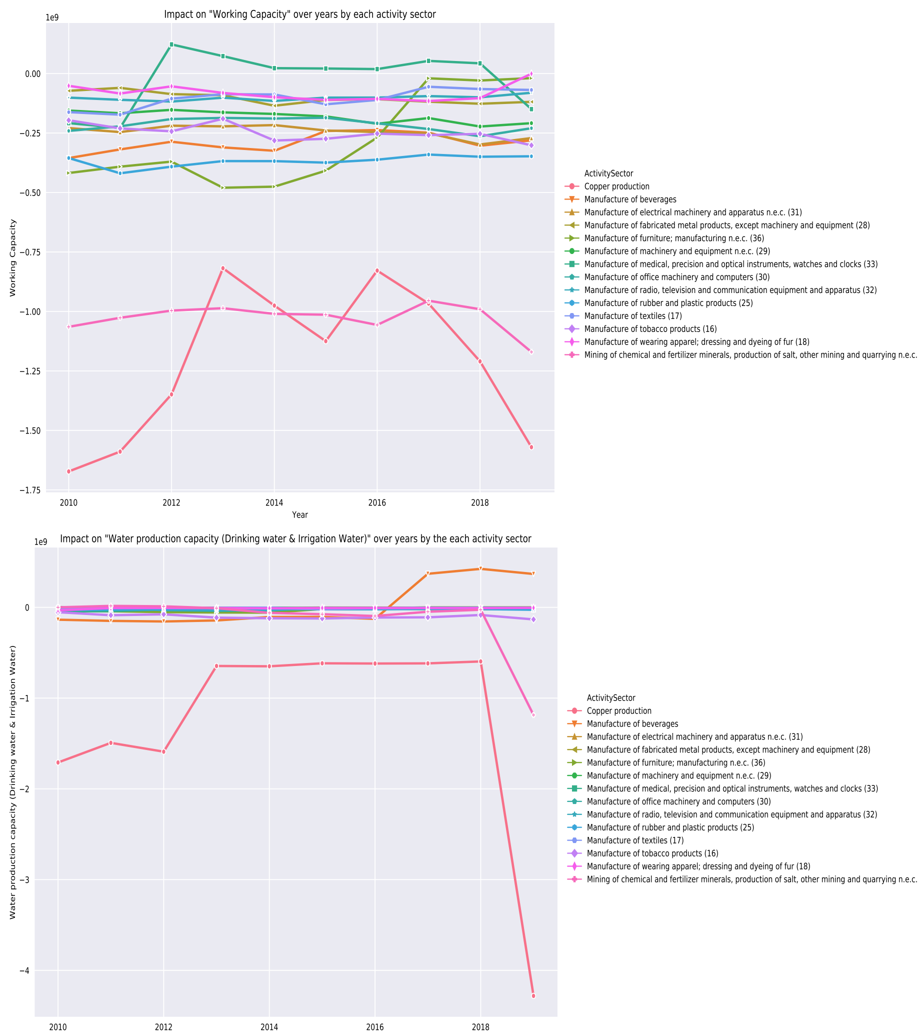

Fig. 7 Manufacturing enterprises impacts on safeguard subjects

production capacity can be adapted and implemented using the proposed approach. In Figs. 7 and 8, negative values indicate negative impacts in terms of impacts monetization, whereas positive values show positive impacts. Moreover, copper production is the most sector that impacts the four selected safeguard subjects. While, in addition to the copper production sector, mining of chemical and fertilizer minerals, production of salt, and other mining and quarrying are impacting working capacity and biodiversity significantly over the years.

Predictive and prescriptive analytics

In the IWAP, the emissions' impacts in terms of the 17 United Nations SDGs were measured based on the different emissions that each enterprise emitted during its operations and procedures. The IWAP team confirmed that the impacts on safeguard subjects that are critical to human well-being and the environment will significantly have impacts on the 17 SDGs. Thus, they gave measurements of these impacts based on the EPS monetization method. Consequently, Fig. 9 shows the correlation matrix between

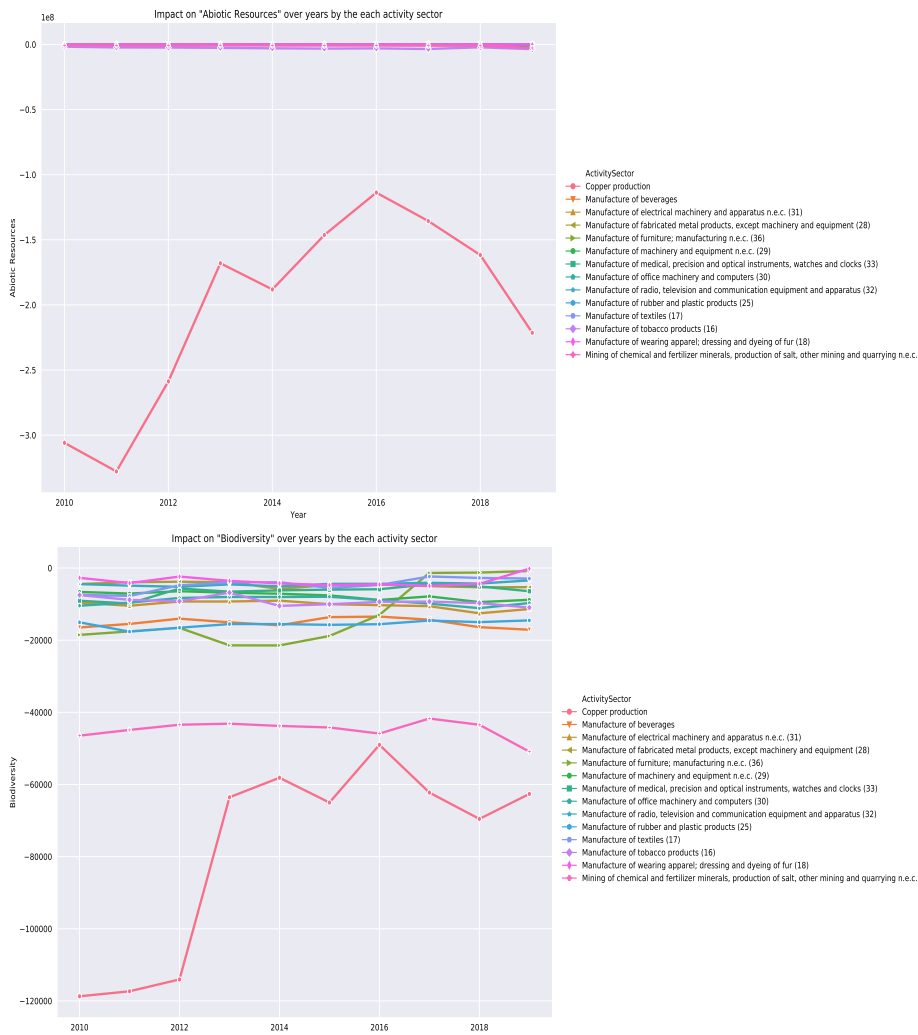

Fig. 8 Manufacturing enterprises impacts on safeguard subjects

the variable within the selected sample of manufacturing enterprises. For instance, the correlation factors between some safeguard subjects and SDGs are equal to 1, such as WC, FishPC, CropPC, MeatPC, SDG 1, SDG 2, SDG 3, SDG 4, SDG 5, SDG 6, SDG 13, SDG 17. These strong correlations can be explained by: the impact on safeguard subjects will result in economic, environmental, and social issues and disasters, such as poverty, climate change, starvation, and lack of agricultural lands. Appendix 3 illustrates every SDG in detail. As a result, based on these correlations, the impact monetization of each SDG based on the correlated safeguard subjects' impacts can be predicted. As part of the developed approach, appropriate dependent variables (X) are assigned to each SDG as a target variable (Y). This enables us to predict and quantify the monetization impacts associated with each SDG. By linking specific dependent variables to the respective SDGs, the developed approach can assess and analyze the financial implications and economic value associated with achieving these sustainable development targets. Finally, the approach facilitates a comprehensive understanding of the monetary aspects and benefits associated with the pursuit of SDGs.

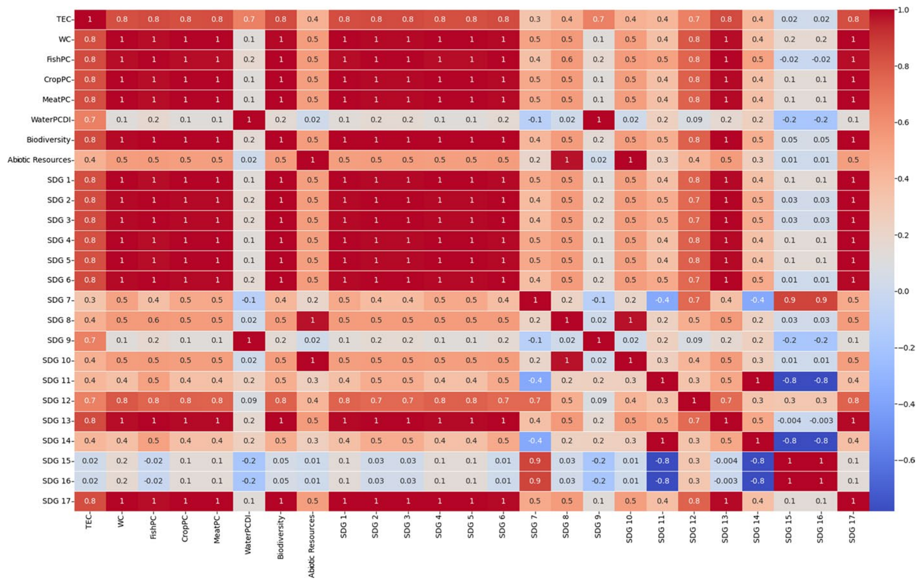

Fig. 9 Correlation matrix between variables

In this part of the predictive analytics, MLlib Library for Apache Spark is used. MLlib is Spark's open-source distributed ML library, as discussed in "Upper-level libraries" section. It provides a number of fundamental statistics, optimization, and linear algebra primitives which give Apache Spark effective functionality for a variety of learning scenarios. Besides, these learning scenarios are supported by a number of programming languages and high-level APIs using Apache Spark robust ecosystem. Hence, these functionalities simplify the use of Apache Spark to create end-to-end ML algorithms and pipelines. In this research work, two ML algorithms are adopted, namely, multiple linear regression (MLR) [1] and artificial neural network (ANN) [15]. Moreover, to evaluate the performance of each algorithm, specific performance metrics are used. These metrics are: mean Squared Error (MSE) and R^2 or coefficient of determination. Where, MSE is the loss function minimization between target SDG y and predicted SDG $\hat{y} = \hat{f}(x)$, given by Eq. 1.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2 \tag{1}$$

And R^2 given by Eq. 2 is a measure used to assess the model's efficacy and validity. Where $\hat{y}_i$ is the estimated value of the dependent value for the ith observation using the model and $\bar{y}$ is the mean of all the observations of the dependent values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}^2)} \tag{2}$$

After evaluating both models on each of the 17 SDGs, Table 2 shows the obtained results based on the above two performance metrics. The performance of both models is significantly accurate which means that both models are performing well on the validation data (impacts monetization data) for each SDG. This significant accuracy of prediction

Table 2 Performance of machine learning models on the 17 SDGs

SDGs (*Y*)	Multiple linear regression			Artificial neural network		
	MSE	R^2	Epochs	MSE	R^2	Epochs
SDG1	1.12067e+07	0.985611	10	0.21056e+09	0.994	100
SDG2	3.01021e+06	0.99459	10	2.33940e+06	0.990334	100
SDG3	3.68953e+08	0.99698	10	2.22087e+08	0.998938	100
SDG4	671584	0.997372	10	0.33027e+05	0.999321	100
SDG5	1.94250e+06	0.980005	10	1.78067e+07	0.9902	100
SDG6	5.8932e+06	0.990303	10	3.98393e+06	0.9987	100
SDG7	2.209393e+07	0.704473	10	1.03928e+08	0.765021	100
SDG8	7.890493e+05	0.92933	10	5.435305e+04	0.90392	100
SDG9	2.89754e+07	0.988229	10	1.45303e+06	0.993241	100
SDG10	1.20943e+09	0.97775	10	1.02343e+08	0.985712	100
SDG11	3029.33	0.969286	10	1.29403e+02	0.97789	100
SDG12	10431.3	0.547476	10	934.55	0.609831	100
SDG13	95848	0.950487	10	10394.5	0.975493	100
SDG14	394.67	0.982938	10	123.43	0.995283	100
SDG15	39283.32	0.97327	10	29182.9	0.985283	100
SDG16	428399	0.962839	10	293839	0.9709	100
SDG17	2.249403e+06	0.959283	10	1.19393e+05	0.9638294	100

(that exceeds 95% for most of the SDGs) demonstrates that the impacts monetization given by IWAP are highly accurate and well measured.

On one hand, this explains that the IWAP calculation methodology of the impacts on safeguard subjects (critical for human health and well-being), as well as, on SDGs is efficient and applicable. On the other hand, this proves that there are strong relationships between the impacts on safeguard subjects and SDGs. More specifically, manufacturing enterprises that have serious impacts on some or all the safeguard subjects will indirectly impact some or all SDGs over the years. For instance, if one of these four safeguard subjects, namely human health (working capacity), crop production capacity, meat production capacity, and fish production capacity significantly impacted by manufacturing enterprises will result in starvation, human health issues, decreasing agricultural productivity, lacking sustainable food production systems, all sorts of epidemics, to name a few. Moreover, this will make these manufacturing enterprises have poor plans to achieve "the 2030 agenda for sustainable development" which includes planning actions for people, the planet, and the prosperity of nations [66]. In addition to economic challenges, the current business that does not envisage and consider environmental, and social challenges in order to achieve the sustainable development goals (SDGs) will put its reputation in danger and will be exceeded by its competitors in the global markets. Finally, the following statements can be highlighted:

1. Integrating life-cycle evaluation of product impacts during a manufacturing enterprise's processes and operations can be considered as its innovation process. Thus, environmental and social impacts are always challenging to measure and evaluate.
2. Impacts monetization coupled with BBA techniques will provide manufacturing enterprises with an efficient life-cycle thinking that supports the decision-making process in sustainable product development.

3. Coupling life cycle sustainability assessment with BDA powerful techniques will allow manufacturing enterprises to analyze their impacts on human health and well-being.
4. Coupling life cycle sustainability assessment with BDA powerful techniques will permit more efficient decision-making on the progress toward the 17 SDGs in order to achieve the 2030 sustainable development agenda.
5. Descriptive and diagnostic analytics based on impact monetization will enable manufacturing enterprises to evaluate and analyze their impacts on the people, and the planet.
6. Predictive analytics will provide insights and information to manufacturing enterprises in order to evaluate their sustainable development.
7. The accuracy and reliability of the numerical results obtained through the developed approach are primarily dependent on the availability and precision of data provided by manufacturing enterprises from various countries. However, it should be noted that the quality of the results may be limited by data availability and the level of detail shared by these enterprises. Additionally, certain manufacturing enterprises may not disclose their environmental and social impacts to the public due to government restrictions or other reasons, further affecting the comprehensiveness of the data.

Conclusion

This paper provided a study of BDA applicability in enhancing manufacturing enterprises' life cycle sustainability assessment (LCSA). It started by giving what LCSA means to manufacturing enterprises in designing and making sustainable products by integrating environmental and social thinking into their business models. Then, it discussed some methods of how the environmental and social impacts can be measured in order to be evaluated and optimized. Thus, based on the EPS method for impacts monetization, a real database that provides impacts measurements of many enterprises around the world was chosen to validate the proposed approach. These impacts measurement concern the impacts monetization on the safeguard subjects as defined by the EPS method during the life cycle of products from their design to the end use (from the cradle to grave). As well as these measurements concern the monetization of impacts in terms of the 17 SDGs as defined by IWP. Moreover, the BDA part of the proposed approach considered two levels of BDA (i.e. descriptive and diagnostic analytics) to analyze manufacturing enterprises' impacts on the safeguard subjects that are critical to human health and well-being. In addition, the two other levels (i.e. predictive and prescriptive analytics) were used to evaluate manufacturing enterprises' progress toward the 2030 sustainable development goals agenda launched by the United Nations. The analysis conducted in this study leads to the following conclusions.

- Sustainable manufacturing enterprises are enhanced and are better positioned to create high returns at a level of risk that is well-balanced when sustainability concepts are incorporated into the investing process.
- Measuring enterprises' impacts on the sustainable environment and society is a hard task.

- Standardizing sustainability metrics is hard because it's difficult to decide which impacts on the environment and society are most severe so to be prioritized in order to reduce the overall environmental and societal impacts.
- Monetization of impacts is one of the best solutions to measure the manifold impacts on the environment and society.
- Practical implications from manufacturing enterprise owners and managers are necessary to allow efficient applicability of the approach in real-life scenarios.

Theoretically, the proposed approach in this study allows scholars to understand how BDA coupled with LCSA thinking can efficiently evaluate the manufacturing enterprises' impacts on the environment and society and make the best decisions toward sustainable development goals. Operationally, the findings presented in the results and analyses of the experiment confirmed the efficiency and applicability of the developed approach. Moreover, they showed that BDA is a promising technology for enhancing manufacturing enterprises' SVC. Finally, in the present study, the analysis was restricted to manufacturing enterprises only and the use of one of the LCSA methods. Thus, as future works, the study will be expanded to more industrial sectors and integrate more sustainable development goals. In addition, since the state of the art is very rich with BDA powerful analytics tools, the study will integrate more tools and techniques within the proposed approach in the near future.

Appendices

Appendix 1: EPS safeguard subjects

See Table 3.

Table 3 The impact categories in the EPS method

Safeguard subject	Impact Category	Indicator name	Unit
Human health	Life Expectancy	YOLL: years of lost life	Person year
	Severe Morbidity & suffering	Severe Morbidity	Person year
	Severe Nuisance	Severe Nuisance	Person year
	Working capacity	Working capacity	Person year
	Migration	Migration	Persons
Ecosystem production capacity	Crop production capacity	Crop	kg
	Wood production capacity	Wood	kg
	Fish & meat production capacity	Fish & meat	kg
	Soil acidification	Base cat-ion capacity	Mole H+ equivalents
	Production capacity of water	Irrigation water	
		Drinking water	kg
			kg
Abiotic stock resource	Depletion of element reserves	Natural gas reserves	kg
		Oil reserves	kg
		Coal reserves	kg
		"Mineral name" reserves	kg of minirals
Biodiversity	Threat contribution: Extinction of a species (animals, plants, organisms and their genes included)	"Normalized extinction of species" and is measured as the share of all red-listed species.	Dimensionless

Cultural and recreational values are very hard to describe with general impact categories and indicators, therefore they are identified only on a case study basis. For instance, culture is very difficult to measure but essential for sustainable development and important to consider. Education, culture consumption, and free press are chosen as state indicators of culture. At present, culture consumption cannot be quantified and is only brought up at a qualitative level.

Appendix 2: Pathways (endpoint effects) on safeguard subjects

See Table 4.

Table 4 Pathways or endpoint effects on safeguard subjects that are resulted from CO_2 emissions to air

Safeguard subject	Substance	Impact category	Category indicator	Pathway (endpoint effects)
Human health	CO_2	Life expectancy	YOLL: years of lost life	Heat stress
				Starvation
				Flooding
				Malaria
				Cold moderation
		Severe morbidity & suffering	Severe morbidity	Starvation
				Malaria
				Heat stress
		Working capacity	Working capacity	Heat stress
		Severe morbidity & suffering	Severe morbidity	Starvation
				Malaria
Ecosystem production capacity	CO_2	Crop production capacity	Crop	Desertification
				climate change
				Rise of sea level
		Fish & meat production capacity	Fish & meat	Starvation
				Desertification
				Draught
		Wood production capacity	Wood	Global warming
				CO_2 fertilization
				Climate change
		Water production capacity	Irrigation water	Climate change
			Drinking water	Climate change
Biodiversity	CO_2	Extinction of a species	Normalized extinction of species (NEX)	Climate change
				Habitat change

Appendix 3: SDGs definitions

See Table 5.

Table 5 SDGs definitions form Serafeim et al. [50]

SDG	Definition
SDG1	By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social, and environmental shocks and disasters
SDG2	By 2030, end hunger and ensure access by all people, particularly the poor and people in vulnerable situations, including infants, to safe, nutritious, and sufficient food all year round
SDG3	By 2030, end all forms of malnutrition, including achieving, by 2025, the internationally agreed targets on stunting and wasting in children under five years of age, and address the nutritional needs of adolescent girls, pregnant and lactating women, and older persons
SDG4	By 2030, double the agricultural productivity and incomes of small-scale food producers, in particular women, indigenous peoples, family farmers, pastoralists, and fishers, including through secure and equal access to land, other productive resources, and inputs, knowledge, financial services, markets and opportunities for value addition and non-farm employment
SDG5	By 2030, ensure sustainable food production systems and implement resilient agricultural practices that increase productivity and production, help maintain ecosystems, strengthen capacity for adaptation to climate change, extreme weather, drought, flooding, and other disasters, and progressively improve land and soil quality
SDG6	By 2030, end the epidemics of AIDS, tuberculosis, malaria, and neglected tropical diseases and combat hepatitis, water-borne diseases, and other communicable diseases
SDG7	By 2030, reduce by one-third premature mortality from non-communicable diseases through prevention and treatment and promote mental health and well-being
SDG8	By 2030, substantially reduce the number of deaths and illnesses from hazardous chemicals and air, water, and soil pollution and contamination
SDG9	Ensure availability and sustainable management of water and sanitation for all
SDG10	By 2030, achieve sustainable management and efficient use of natural resources
SDG11	By 2025, prevent and significantly reduce marine pollution of all kinds, particularly from land-based activities, including marine debris and nutrient pollution
SDG12	By 2020, sustainably manage and protect marine and coastal ecosystems to avoid significant adverse impacts, including strengthening their resilience, and take action for their restoration to achieve healthy and productive oceans
SDG13	Minimize and address the impacts of ocean acidification, including through enhanced scientific cooperation at all levels
SDG14	Enhance the conservation and sustainable use of oceans and their resources by implementing international law as reflected in UNCLOS, which provides the legal framework for conserving and sustainable use of oceans and their resources
SDG15	By 2020, ensure the conservation, restoration, and sustainable use of terrestrial and inland freshwater ecosystems and their services, particularly forests, wetlands, mountains, and drylands, in line with obligations under international agreements
SDG16	By 2020, promote the implementation of sustainable management of all types of forests, halt deforestation, restore degraded forests and substantially increase afforestation and reforestation globally
SDG17	Take urgent and significant action to reduce the degradation of natural habitats, halt the loss of biodiversity and, by 2020, protect and prevent the extinction of threatened species

Acknowledgements

The authors gratefully acknowledge the financial support by the Eiffel Excellence Scholarship Program of Campus France under the reference 116707T.

Author contributions

Mr. LT: Paper writing, problem formulation, approaches proposal and experimental performing. Prof. LB: Paper writing, problem formulation, approaches proposal and experimental performing. Prof. ANSM: Problem formulation and approaches proposal. Prof. MDEO: Problem formulation and approaches proposal.

Funding

This work was supported by the Eiffel Excellence Scholarship Program of Campus France and Aix-Marseille University, France.

Availibility of data and materials
The used data and materials are available when requested.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
The authors give the Publisher permission to publish the work.

Competing interests
The authors declare that they have no competing interests.

Received: 24 May 2023 Accepted: 17 October 2023
Published online: 11 November 2023

References

1. Adams Hayes. Multiple linear regression (mlr) definition, formula, and example. https://www.investopedia.com/terms/m/mlr.asp. accessed 28 2022.
2. Afrin K, Iquebal AS, Kumar SK, Tiwari M, Benyoucef L, Dolgui A. Towards green automated production line with rotary transfer and turrets: a multi-objective approach using a binary scatter tabu search procedure. International Journal of Computer Integrated Manufacturing. 2016;29:768–85. https://doi.org/10.1080/0951192X.2015.1109143.
3. Ahlroth S, Developing a weighting set based on monetary damage estimates: Method and case studies. KTH Royal Institute of Technology 2009.
4. An X, Si G, Xia T, Liu Q, Li Y, Miao R. Operation and maintenance optimization for manufacturing systems with energy management. Energies. 2022;15. https://doi.org/10.3390/en15197338.
5. Arendt R, Bachmann TM, Motoshita M, Bach V, Finkbeiner M. Comparison of different monetization methods in lca: A review. Sustainability. 2020;12. https://doi.org/10.3390/su122410493.
6. Armbrust M, Xin RS, Lian C, Huai Y, Liu D, Bradley JK, Meng X, Kaftan T, Franklin MJ, Ghodsi A, Zaharia M. Spark sql: Relational data processing in spark, in: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, Association for Computing Machinery. 2015;1383-1394. https://doi.org/10.1145/2723372.2742797.
7. Arunachalam D, Kumar N, Kawalek JP. Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. Transportation Research Part E: Logistics and Transportation Review. 2018;114:416–36.
8. Battaïa O, Benyoucef L, Delorme X, Dolgui A, Thevenin S. Sustainable and Energy Efficient Reconfigurable Manufacturing Systems. 2020;179–191. https://doi.org/10.1007/978-3-030-28782-5_9.
9. Bengt S, Sara P. A selection of safeguard subjects and state indicators for sustainability assessments. The International Journal of Life Cycle Assessment 21 2016.
10. Bibri S, Alahi A, Sharifi A, Krogstie J. Environmentally sustainable smart cities and their converging ai, iot, and big data technologies and solutions: an integrated approach to an extensive literature review. Energy Informatics. 2023;6:32. https://doi.org/10.1186/s42162-023-00259-2.
11. Boissier M, Krueger J, Wust J, Plattner H. An integrated data management for enterprise systems. 2014. https://doi.org/10.5220/0004897204100418.
12. Camarinha-matos LM, Afsarmanesh H. Collaborative networks: Value creation in a knowledge society. IFIP International Federation for Information Processing. 2006;207:26–40.
13. Centre JR, for Prospective Technological Studies, I., Weidema, B., Environmental improvement potentials of meat and dairy products. Publications Office. https://doi.org/10.2791/38863 2008.
14. Dahmani N, Benhida K, Belhadi A, Kamble S, Elfezazi S, Jauhar SK. Smart circular product design strategies towards eco-effective production systems: A lean eco-design industry 4.0 framework. Journal of Cleaner Production 320, 2021;128847. https://doi.org/10.1016/j.jclepro.2021.128847.
15. Dastres R, Soori M. Artificial neural network systems. International Journal of Imaging and Robotics. 2021;21:13–25.
16. Eldh P, Johansson J. Weighting in lca based on ecotaxes - development of a midpoint method and experiences from case studies. International Journal of Life Cycle Assessment - INT J LIFE CYCLE ASSESS. 2006;11:81–8. https://doi.org/10.1065/lca2006.04.015.
17. Eslami Y, Lezoche M, Panetto H, Dassisti M. On analysing sustainability assessment in manufacturing organisations: a survey. International Journal of Production Research. 2021;59:4108–39. https://doi.org/10.1080/00207543.2020.1755066.
18. Evans S, Gregory M, Ryan C, Bergendahl M, Tan A. Towards a sustainable industrial system: With recommendations for education, research, industry and policy. Institute for Manufacturing: University of Cambridge; 2009.
19. Ferrari AM, Volpi L, Settembre-Blundo D, García-Muiña FE. Dynamic life cycle assessment (lca) integrating life cycle inventory (lci) and enterprise resource planning (erp) in an industry 4.0 environment. Journal of Cleaner Production 286, 2021;125314. https://doi.org/10.1016/j.jclepro.2020.125314.
20. Fosso Wamba S, Akter S. Big data analytics for supply chain management: A literature review and research agenda, 2015;61–72. https://doi.org/10.1007/978-3-319-24626-0_5.

21. Fosso Wamba S, Akter S, Edwards A, Chopin G, Gnanzou D. How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics. 2015;165:234–46. https://doi.org/10.1016/j.ijpe.2014.12.031.
22. Freiberg D, Park D, Serafeim G, Zochowski TR. Corporate environmental impact: Data supplement. 2023. https://www.hbs.edu/impact-weighted-accounts/Pages/research.aspx?topic=Data.
23. Gao RX, Wang P. Sensors to control processing and improve lifetime and performance for sustainable manufacturing. In: Abraham MA, editors. Encyclopedia of sustainable technologies, Elsevier; 2017. p. 447–62. https://doi.org/10.1016/B978-0-12-409548-9.10217-9.
24. Guinée J. Life cycle sustainability assessment: what is it and what are its challenges?. In: Clift R, Druckman A, editors. Taking Stock of Industrial Ecology. Cham: Springer; 2016. https://doi.org/10.1007/978-3-319-20571-7_3.
25. Gupta SM, Güngör A, Govindan K, Özceylan E, Kalaycı CB, Piplani R. Responsible & sustainable manufacturing. International Journal of Production Research. 2020;58:7181–2. https://doi.org/10.1080/00207543.2020.1841968.
26. Hannouf M, Assefa G, Hannouf M, Gates I. Cause-effect chains in s-lca based on dpsir framework using markov healthcare model: an application to "working hours" in canada. The International Journal of Life Cycle Assessment. 2021;26. https://doi.org/10.1007/s11367-021-01900-6.
27. Hu H, Wen Y, Chua TS, Li X. Toward scalable systems for big data analytics: A technology tutorial. IEEE Access. 2014;2:652–87. https://doi.org/10.1109/ACCESS.2014.2332453.
28. Igor De Souza, 2019. What is apache spark? http://www.igfasouza.com/blog/what-is-apache-spark/. accessed 28 (2022).
29. Jamwal A, Agrawal R, Sharma M, Giallanza A. Industry 4.0 technologies for manufacturing sustainability: A systematic review and future research directions. Applied Sciences 11. 2021. https://doi.org/10.3390/app11125725.
30. Karau H, Konwinski A, Wendell P, Zaharia M. Learning spark: lightning-fast big data analysis. "O'Reilly Media, Inc." 2015.
31. Kiel D, Arnold C, Müller J, Voigt KI. Sustainable industrial value creation: Benefits and challenges of industry 4.0. International Journal of Innovation Management 21, 2017;1740015.
32. Kim H, Park J, Jang J, Yoon S. Deepspark: Spark-based deep learning supporting asynchronous updates and caffe compatibility arXiv:abs/1602.08191 2016.
33. Klöpffer W. Life cycle sustainability assessment of products. The International Journal of Life Cycle Assessment. 2008;13:89–95. https://doi.org/10.1065/lca2008.02.376.
34. Küster T, Lützenberger M, Freund D, Albayrak S, DAI-Labor. Distributed evolutionary optimisation for electricity price responsive manufacturing using multi-agent system technology, 2013.
35. Lake A, Acquaye A, Genovese A, Kumar N, Koh S. An application of hybrid life cycle assessment as a decision support framework for green supply chains. International Journal of Production Research. 2015;53:6495–521. https://doi.org/10.1080/00207543.2014.951092.
36. Lv Z, Iqbal R, Chang V. Big data analytics for sustainability. Future Generation Computer Systems. 2018;86:1238–41.
37. Mahasivam N, Nikolov N, Sukhobok D, Roman D. Data preparation as a service based on apache spark, 2017;125–139. https://doi.org/10.1007/978-3-319-67262-5_10.
38. Majeed A, Zhang Y, Ren S, Lv J, Peng T, Waqar S, Yin E. A big data-driven framework for sustainable and smart additive manufacturing. Robotics and Computer-Integrated Manufacturing. 2021;67:102026.
39. Martin J, Petty W, Wallace J. Shareholder value maximization-is there a role for corporate social responsibility? Journal of Applied Corporate Finance. 2009;21:110–8.
40. Martí C, Anna F. The societal impact of big data: A research roadmap for europe. Technology in Society. 2018;54:74–86.
41. Meng X, Bradley JK, Yavuz B, Sparks ER, Venkataraman S, Liu D, Freeman J, Tsai DB, Amde M, Owen S, Xin D, Xin R, Franklin MJ, Zadeh R, Zaharia M, Talwalkar A. Mllib: Machine learning in apache spark. Journal of Machine Learning Research 2015. arXiv:abs/1505.06807.
42. Mirarab A, Mirtaheri S, Asghari SA. Value creation with big data analytics for enterprises: a survey. TELKOMNIKA Indonesian Journal of Electrical Engineering. 2019;17:2790–802.
43. Mouzon G, Yildirim MB, Twomey J. Operational methods for minimization of energy consumption of manufacturing equipment. International Journal of Production Research. 2007;45:4247–71. https://doi.org/10.1080/00207540701450013.
44. Nguyen T, ZHOU L, Spiegler V, Ieromonachou P, Lin Y. Big data analytics in supply chain management: A state-of-the-art literature review. Computers & Operations Research. 2018;98:254–64. https://doi.org/10.1016/j.cor.2017.07.004.
45. Niu Y, Jiang Z, Geng N, Jiang S. Disclosing the formation and value creation of servitization through influential factors: a systematic review and future research agenda. International Journal of Production Research. 2020;59:1–33. https://doi.org/10.1080/00207543.2020.1834158.
46. Qasim Shabbir M, Gardezi S. Application of big data analytics and organizational performance: the mediating role of knowledge management practices. Journal of Big Data. 2020;7. https://doi.org/10.1186/s40537-020-00317-6.
47. Rejeb A, Suhaiza Z, Rejeb K, Seuring S, Treiblmaier H. The internet of things and the circular economy: A systematic literature review and research agenda. Journal of Cleaner Production. 2022;350:131439. https://doi.org/10.1016/j.jclepro.2022.131439.
48. Sala S, Vasta A, Mancini L, Dewulf J. Rosenbaum E. Social life cycle assessment: State of the art and challenges for supporting product policies. Publications Office of the European Union 2015.
49. Salloum S, Dautov R, Chen X, Peng P, Huang J. Big data analytics on apache spark. International Journal of Data Science and Analytics. 2016;1. https://doi.org/10.1007/s41060-016-0027-9.
50. Serafeim G, Park D, Freiberg D, Zochowski R. Corporate environmental impact: Measurement, data and information. SSRN Electronic Journal. 2020. https://doi.org/10.2139/ssrn.3565533.
51. Shoro A, Soomro T. Big data analysis: apache spark perspective. Glob J Comput Sci Technol. 2015;15.
52. Singh D, Reddy C. A survey on platforms for big data analytics. Journal of Big Data. 2014;2. https://doi.org/10.1186/s40537-014-0008-6.

53. Soori M, Arezoo B, Dastres R. Internet of things for smart factories in industry 4.0, a review. Internet of Things and Cyber-Physical Systems. 2023;3:192–204. https://doi.org/10.1016/j.iotcps.2023.04.006.
54. Souza GC. Supply chain analytics. Business Horizons. 2014;57:595–605.
55. Steen B. A systematic approach to environmental priority strategies in product development (eps).Version 2000–General system characteristics. Gothenburg Centre Environ Assess Prod Material Syst. 1999.
56. Steen B. Calculation of monetary values of environmental impacts from emissions and resource use the case of using the eps 2015d impact assessment method. Journal of Sustainable Development. 2016;9:15.
57. Steen, B. Monetary Valuation of Environmental Impacts: Models and Data (1st ed.). CRC Press; 2019. https://doi.org/10.1201/9780429430237.
58. Tamym L, Benyoucef L, Moh ANS, Ouadghiri MDE. Sustainable value creation of networked manufacturing enterprises: Big data analytics based methodology. IFAC-PapersOnLine. 2022;55:804–9. https://doi.org/10.1016/j.ifacol.2022.09.508.
59. Tamym L, Benyoucef L, Moh Nait Sidi A, El Ouadghiri MD. A big data based architecture for collaborative networks: Supply chains mixed-network. Computer Communications. 2021;175:102–11. https://doi.org/10.1016/j.comcom.2021.05.008.
60. Tamym L, Benyoucef L, Moh Nait Sidi, A, El Ouadghiri MD. Big data analytics-based approach for robust, flexible and sustainable collaborative networked enterprises. Advanced Engineering Informatics. 2023;55: 101873. https://doi.org/10.1016/j.aei.2023.101873.
61. Tamym L, Benyoucef L, Nait Sidi Moh A, El Ouadghiri MD. Towards big data-based sustainable business models and sustainable supply chain, in: Emerging Trends in Intelligent Systems and Network Security, Springer International Publishing. 2023b;464–474. https://doi.org/10.1007/978-3-031-15191-0_44.
62. Tamym L, Nait Sidi Moh A, Benyoucef L, El Ouadghiri MD. Goods and activities tracking through supply chain network using machine learning models. In: Dolgui A, Bernard A, Lemoine D, von Cieminski G, Romero D editors. Advances in production management systems. Artificial intelligence for sustainable and resilient production systems. APMS 2021. IFIP Advances in information and communication technology, vol 630. Cham: Springer; 2021. https://doi.org/10.1007/978-3-030-85874-2_1.
63. Thakker V, Bakshi BR. Toward sustainable circular economies: A computational framework for assessment and design. Journal of Cleaner Production. 2021;295:126353. https://doi.org/10.1016/j.jclepro.2021.126353.
64. Tomas R, Bengt S. Environmental priority strategies in product design (eps). https://www.lifecyclecenter.se/projects/environmental-priority-strategies-in-product-design-eps/. accessed 28 Oct 2022 2015.
65. Tsai WH, Tsaur TS, Chou YW, Liu JY, Hsu JL, Hsieh CL. Integrating the activity-based costing system and life-cycle assessment into green decision-making. International Journal of Production Research. 2015;53:451–65. https://doi.org/10.1080/00207543.2014.951089.
66. United Nations. Transforming our world: the 2030 agenda for sustainable development. https://sdgs.un.org/2030agenda. accessed 25 Nov 2022 2015.
67. United Nations. Big data for sustainable development. https://www.un.org/en/global-issues/big-data-for-sustainable-development. accessed 25 Nov 2022; 2022.
68. Vogtländer J, Peck D, Kurowicka D. The eco-costs of material scarcity, a resource indicator for lca, derived from a statistical analysis on excessive price peaks. Sustainability. 2019;11. https://doi.org/10.3390/su11082446.
69. Xin R, Gonzalez JE, Franklin MJ, Stoica I. Graphx: a resilient distributed graph system on spark. First International Workshop on Graph Data Management Experiences and Systems; 2013.
70. Yamaguchi K, Ii R, Itsubo N. Ecosystem damage assessment of land transformation using species loss. The International Journal of Life Cycle Assessment. 2016;23:2327–38.
71. Yang M, Vladimirova D, Rana P, Evans S. Sustainable value analysis tool for value creation. Asian Journal of Management Science and Applications. 2014;1:312–32.
72. Ye L, Pan SL, Wang J, Wu J, Dong X. Big data analytics for sustainable cities: An information triangulation study of hazardous materials transportation. Journal of Business Research. 2021;128:381–90. https://doi.org/10.1016/j.jbusres.2021.01.057.
73. Zaharia M. Models Built over RDDs. An architecture for fast and general data processing on large clusters. Assoc Comput Mach Morgan Claypool. 2018. https://doi.org/10.1145/2886107.2886111.
74. Zhang D, Pan SL, Yu J, Liu W. Orchestrating big data analytics capability for sustainability: A study of air pollution management in china. Information & Management. 2019;103231. https://doi.org/10.1016/j.im.2019.103231.
75. Zhou J, Xiahou T, Liu Y. Multi-objective optimization-based topsis method for sustainable product design under epistemic uncertainty. Applied Soft Computing. 2021;98:106850. https://doi.org/10.1016/j.asoc.2020.106850.
76. Ziegler P, Dittrich KR. Data Integration — Problems, Approaches, and Perspectives. Springer Berlin Heidelberg. 2007;39–58. https://doi.org/10.1007/978-3-540-72677-7_3.

Publisher's Note