technologies

Review

Bridging the Maturity Gaps in Industrial Data Science: Navigating Challenges in IoT-Driven Manufacturing

Amruta Awasthi , Lenka Krpalkova and Joseph Walsh *

School of Science, Technology, Engineering and Mathematics (STEM), Munster Technological University (MTU), Kerry Campus, V92 HD4V Kerry, Ireland; amruta.awasthi@research.ittralee.ie (A.A.); lenka.krpalkova@mtu.ie (L.K.)
* Correspondence: joseph.walsh@mtu.ie

Abstract: This narrative review evaluates the curtail components of data maturity in manufacturing industries, the associated challenges, and the application of industrial data science (IDS) to improve organisational decision-making. As data availability grows larger, manufacturing organisations face difficulties comprehending heterogeneous datasets of varying quality, which may lead to inefficient decision-making and other operational inefficiencies. It underlines that data appropriate for its intended application is considered quality data. The importance of including stakeholders in the data review process to enhance the data quality is accentuated in this paper, specifically when big data analysis is to be integrated into corporate strategies. Manufacturing industries leveraging their data thoughtfully can optimise efficiency and facilitate informed and productive decision-making by establishing a robust technical infrastructure and developing intuitive platforms and solutions. This study highlights the significance of IDS in revolutionising manufacturing sectors within the framework of Industry 4.0 and the Industrial Internet of Things (IIoT), demonstrating that big data can substantially improve efficiency, reduce costs, and guide strategic decision-making. The gaps or maturity levels among industries show a substantial discrepancy in this analysis, which is classified into three types: Industry 4.0 maturity levels, data maturity or readiness condition index, and industrial data science and analytics maturity. The emphasis is given to the pressing need for resilient data science frameworks enabling organisations to evaluate their digital readiness and execute their data-driven plans efficiently and effortlessly. Simultaneously, future work will focus on pragmatic applications to enhance industrial competitiveness within the heavy machinery sector.

Keywords: data-driven transformation; heavy industries; maturity levels; industrial gaps; data analysis/analytics; predictive analytics/maintenance; artificial intelligence (AI); machine learning (ML); deep learning (DL); Industry 4.0

Academic Editor: Eugene Wong and Lipo Wang

Received: 21 October 2024
Revised: 18 December 2024
Accepted: 25 December 2024
Published: 6 January 2025

Citation: Awasthi, A.; Krpalkova, L.; Walsh, J. Bridging the Maturity Gaps in Industrial Data Science: Navigating Challenges in IoT-Driven Manufacturing. *Technologies* **2025**, *13*, 22. https://doi.org/10.3390/technologies13010022

1. Introduction

Manufacturing industries are a traditional global industry that is famously conservative and hesitant to adopt innovations, resulting in unique gaps in its growth. However, they have proliferated with the rise of Industry 4.0, the Industrial Internet of Things (IIoT), big data, and other technological advancements. The advancements in these areas have transformed the sectors into digital ecosystems to compete in the current market and make them more data-driven [1]. Rapid technological advancements have led to a substantial change in data collection techniques, especially with the combination of cloud computing and the Internet of Things (IoT). These developments have allowed the accumulation of enormous volumes of data from various sources, such as assets, consumer contacts, and industrial processes. However, a new set of opportunities and difficulties arises with the arrival of cloud solutions, advanced networking capabilities, and intelligent sensors. Furthermore, Gartner indicates that robotics, automation, data analytics, and the Internet of Things (IoT) enhance productivity, predictive accuracy, and connection [2]. Despite the advantages, significant barriers persist.

In the age of Industry 4.0 and sophisticated information technology, large-scale manufacturing processes are perpetually advancing towards grander scale, complexity, and intelligence for their transformation [3]. At the heart of this transformation lies the concept of IDS, which harnesses the vast amounts of data generated by industrial operations. In Industry 4.0, cutting-edge technologies like automation and information technology are causing a significant shift in the industrial sector. Numerous advantages result from this transformation, including increased productivity, better operational efficiency, improved adaptability to market fluctuations, and the capacity to provide consumers with customised and personalised solutions. It optimises resources, such as energy efficiency and reduced wastage. Data-driven industrial systems require industrial data science to extract knowledge and insights using various subjects, means, techniques, and methodologies it offers. Industrial data science helps the industry develop from infancy to full-fledged data-driven with a "whatever it takes" attitude to get there [4]. To fully embrace Industry 4.0, there are challenges to overcome. Businesses seeking to incorporate these cutting-edge technologies into their operations must engage in careful and detailed planning, promote collaboration across different departments, and make significant financial investments to ensure that their efforts align with their strategic goals and ultimately result in a profitable return on investment. To achieve that goal, industrial organisations have entered a new era of "Big Data", in which the volume, velocity, value, veracity, and variety of data they handle are expanding at breakneck speed [5]. Industries that are automated, data-driven, and have implemented Industry 4.0 standards are more profitable than those that have not. This is because the number of interjections caused by humans decreases, and performance increases. Nonetheless, the initial capital expenses for adopting data-driven Industry 4.0 automation are high, and operational constraints such as a lack of infrastructure, insufficient and incomplete data, compartmentalised processes, and difficulty managing exceptions are significant [6].

Data created by sensors incorporated in industrial machinery, network solutions, and business management have already reached a total volume of more than 1000 Exabytes yearly in modern industries, and this figure is projected to rise in the next few years. Industries will be able to improve their performance by obtaining and evaluating data throughout the product lifecycle, thanks to new data-driven methods [7–9].

Against this backdrop, this narrative review addresses critical research questions that explore the maturity gaps in data science capabilities within manufacturing industries. Specifically, this study seeks to understand the following:

1. What are the current maturity levels of data readiness among manufacturing firms, and how do these levels influence their capability to adopt Industry 4.0 technologies?
2. What specific challenges do organisations face in adopting and integrating data-driven practices, particularly regarding collecting, storing, and analysing industrial data?
3. How can organisations bridge the identified maturity gaps to fully leverage the potential of big data and advanced analytics in their operations?

There are several challenges in adopting data-driven technology, especially in the initial investment and infrastructure. The Industry 4.0 transition often requires significant capital investment, which many manufacturing industries hesitate to undertake. Like in other sectors, the fundamental challenge with digitisation in heavy machinery manufacturing is data storage and a lack of data standards. Many operations managers have said unequivocally that the current state of low data quality and limited adoption of industrial data science applications is insufficient to fully automate heavy machinery industries in the future [10–12]. The following are some reasons why it is inadequate and why additional studies are needed in this field [13,14]:

1. The lack of a systematic, intelligible data pool makes it challenging to inspect and analyse equipment operations and performance in a timely manner [15,16].
2. Data standards, formats, and structures may be incompatible or missing, preventing manufacturing industries from successfully collecting and processing data [17,18].

Industrial data science is critical for the development and innovations in heavy machinery manufacturing industries or every industry, as it gives endless possibilities to turn unused and unexplored data into a source of intelligent decision-making and generating revenue, thus leading the industry to betterment, progress, and a competitive edge. This paper highlights the gaps, challenges, and maturity levels that heavy industrial manufacturing faces in becoming data-driven in the disruptive digitalisation era of Industry 4.0. The following sections will explore the complexities of data quality, analytical approaches, and the function of predictive analytics in improving operational performance. It will also tackle the issues associated with deploying robust industrial data science frameworks and provide insights into successful ways for surmounting these limitations. This study seeks to promote an informed dialogue regarding the future of industrial data practices in heavy machinery. This study underscores the importance of a systematic approach to data science for attaining transformative growth and improved competitiveness in the manufacturing industry. The workflow structure is depicted in Figure 1 below.

Figure 1. Research paper workflow.

2. Theoretical Framework

The technological framework for IDS, especially for Industry 4.0 and the Industrial Internet of Things (IIoT), constitutes a multifaceted system comprising several critical components and layers. This system facilitates the seamless integration, comprehensive analysis, and efficient application of data across diverse industrial processes.

The framework fundamentally facilitates data acquisition from various sources, such as sensors, machines, and operational systems, integrated via sophisticated networking technologies. These elements collaborate to deliver instantaneous insights into production efficiency, equipment performance, and supply chain dynamics.

Data analytics is integral to this paradigm. It employs machine learning, predictive analytics, and data visualisation techniques to derive significant patterns and trends from the acquired data. This analysis assists manufacturers in making educated decisions, optimising processes, raising product quality, and improving operational efficiency.

Furthermore, the framework tackles the unique issues faced by the manufacturing industry during its digital transformation, including data maturity levels and deficiencies, concerns regarding data quality and quantity, and the necessity for interoperability among various systems. By integrating best practices in data governance and enhancing maturity levels, the framework guarantees that data utilised in the industrial sector is proficient in decision-making.

The technology framework for IDS is essential for manufacturers to fully leverage their data, fostering innovation, enhancing competitiveness, and facilitating a more agile and responsive production landscape.

2.1. Industrial Data Science

Industrial data science is an emerging field that combines traditional data science techniques with domain-specific knowledge to solve real-world industrial problems [19]. Industry 4.0's digital revolution and 5.0's product customisation have given businesses unprecedented opportunities to extract value from their data [20]. Data have risen to prominence in this rapidly changing environment as primary sources of innovation and business success. If properly handled, the enormous volumes of data that organisations now hold help them improve their operations, cost savings, and decision-making capacities [21,22].

Industrial data science makes a significant contribution to technological advancement. Unsurprisingly, industrial data science heavily relies on data, algorithms, computers, and the proverbial search for a needle in a haystack. It necessitates a strong connection between technical knowledge, business knowledge, and the capacity to use analytical methodologies [23,24].

Consumer and industrial data are vastly different. Data are developed by people who employ data science approaches in various areas. Customers produce data in various ways, be it through purchases, website access, and phone conversations. In contrast, the industrial sector works with machine-generated data. This data is usually generated in a fully automated manner. While having a large amount of data might help Machine Learning (ML), it also comes with its own set of difficulties and limitations [25].

Industrial data science approaches may be applied to all stages of the data science lifecycle, from data preparation to data analysis and modelling to predictive modelling. It aids in defining essential business questions in the context of digitised processes, the discovery and curation of data needed to answer those questions, as well as the building of predictive models that enhance operations, customer happiness, and the industry's bottom line. Industrial data science integrates computational, mathematical, and statistical techniques with production domain expertise across every industry. It includes various subject areas such as from the technical side—data mining, data analysis, data analytics, statistical analysis, artificial intelligence (AI), ML, deep learning (DL), neurocomputing, big data, pattern recognition, mathematics—and from the business side—business intelligence (BI), business analytics, business development [26–28].

Industrial data science can drive efficiency, reduce costs, and foster innovation across various industries. It has wide-ranging applications across multiple sectors. Some of the critical areas where it is making a significant impact are as follows:

- Manufacturing:
    a. Quality control: ML models detect defects in products or processes using computer vision, ensuring higher quality and waste reduction [29].
    b. Predictive maintenance: Industries can predict when a piece of equipment will likely fail by analysing sensor and past maintenance data from machines and performing maintenance before failure occurs, thereby reducing downtime and costs [29].
- Finance:
    a. Fraud detection: ML algorithms can identify unusual transaction patterns and predict fraudulent activities [30].
    b. Risk management: Data science models help investors make smarter investment decisions by evaluating risks and forecasting market trends [30].
- Energy:
    a. Predictive maintenance: Analogous to manufacturing, predictive maintenance in the energy sector aids in maintaining infrastructure and minimising interruptions [31].
    b. Smart grids: Data science enhances the allocation of electricity, regulating supply and demand instantaneously [31].
- Healthcare:
    a. Personalised medicine: Healthcare professionals can deliver personalised treatment regimens based on patient data analysis [32].
    b. Predictive analytics: Data science facilitates the prediction of disease outbreaks, patient readmissions, and treatment outcomes, resulting in enhanced healthcare administration [32].
- Transportation:
    a. Route optimisation: Evaluating traffic patterns and meteorological conditions to identify the most efficient logistics and public transit routes [32].
    b. Autonomous vehicles: Data science is essential in advancing autonomous vehicles by analysing extensive sensor data to facilitate real-time decision-making [32].

This study article aims to clarify the disparities in the heavy machinery manufacturing sector by highlighting the various levels of maturity that multiple industries need to reach to compete successfully with data-driven rivals. This article attempts to thoroughly understand the complexities associated with incorporating data-driven strategies and technologies to maintain competitive advantage by exploring the nuances and challenges faced by various heavy machinery manufacturing industry sectors.

2.2. Developing a Data-Driven Culture in Industries

Industrial engineering has played a crucial role in the success of manufacturing companies for many years. With the emergence of Industry 4.0 and now 5.0 and the increasing importance of data analytics, it is essential to integrate new methods with established industrial engineering practices. Industrial engineering must evolve to incorporate data analytics and IDS to stay relevant in the Industry 4.0 era [24].

Developing and implementing a data strategy is difficult, especially if the company or sector has never had a data-driven culture. Industries need to ascertain their level of maturity or standing in the roadmap to becoming data-driven. Industry maturity refers to "a condition of readiness" for data and big data strategy execution within the industry [33]. Following a data strategy and a data roadmap fit for the business's developmental goals, generally with a multi-disciplinary team, is vital for developing a data-driven culture in a firm [34].

The authors in [32] propose a framework for a data-driven culture in the construction industry comprising five components: production culture, usage culture, data cultivation, datafication, and data infrastructure. The organisation's digital competitiveness is contingent upon four primary drivers of data: data analytics, data literacy, data democratisation, and data leadership.

The five maturity stages for developing a data-driven culture in an organisation are exhibited below [33]:

1. LEVEL 1: Infancy
    - Initial considerations about Big Data
        - Initiates discussions on the relevance and potential of Big Data
        - Identifies key stakeholders and champions within the organisation
        - Evaluate existing data sources and technologies available
    - Basic Big Data environment established
        - Sets up initial data storage solutions or pilot projects
        - Ensures data privacy and compliance measures are considered
        - Conducts a gap analysis to identify the required capabilities
    - Proof-of-concept or pilot projects in progress
        - Runs small-scale experiments to demonstrate value and viability
        - Collects feedback and insights to inform future Big Data strategies
        - Creates documentation to capture lessons learned for future initiatives
2. LEVEL 2: Technical Adoption
    - Using Big Data mainly for storage and transformation
        - Implements data lakes or big data repositories for large datasets
        - Utilises ETL (Extract, Transform, Load) processes for data integration
        - Implements data archiving solutions to optimise storage
    - Primarily utilised by the IT department
        - Focuses on infrastructure and data management rather than end-user analytics
        - Establishes foundational technologies, such as Hadoop or Spark
        - Develops internal dashboards for monitoring data usage and performance
    - Some exploratory analytics conducted using Big Data
        - Conducts initial data exploration projects to identify potential use cases
        - Tests various analytical tools and platforms for feasibility
        - Engages in proof-of-concept projects for different analytical methodologies
3. LEVEL 3: Business Adoption
    - Leveraging discrete use cases for specific lines of business (LOBs)
        - Develops targeted analytical solutions for marketing, sales, or operations
        - Adjusts processes based on findings from data analysis
        - Encourages interdepartmental collaboration on analytics initiatives

- Analysis of both structured and unstructured data
    - Incorporates data from social media, customer feedback, and sensors
    - Utilises advanced analytical techniques, like natural language processing
    - Conducts sentiment analysis to gauge customer opinions
- Application of predictive analytics to Big Data
    - Employs machine learning algorithms to identify trends and patterns
    - Conducts scenario analysis to improve strategic planning
    - Implements A/B testing frameworks to validate predictive models

4. LEVEL 4: Enterprise Adoption
    - Utilising use cases across multiple lines of business (LOBs)
        - Identifies key performance indicators (KPIs) that span departments
        - Implements solutions that drive value across the organisation
        - Conducts regular reviews of use case effectiveness and ROI
    - Integrated metadata, quality, and governance across Big Data
        - Employs automated tools for data quality assessment
        - Creates a centralised repository for metadata management
        - Develops clear data ownership and stewardship policies
    - Predictive insights are integrated into business operations
        - Implements dashboards for real-time monitoring and decision-making
        - Trains staff on interpreting and acting on predictive insights
        - Establishes a feedback loop to refine predictive models based on outcomes
5. LEVEL 5: Data and Analytics as a Service
    - Operating as a "data service provider"
        - Provides APIs for data access and analytics tools
        - Supports third-party applications and services
        - Offers real-time data streaming capabilities
    - Self-service data access
        - Enables users to generate their own reports and dashboards
        - Offers training and resources for data literacy
        - Implements user-friendly interfaces for data exploration
    - Collaboration and sharing of analytics across the enterprise
        - Establishes data governance frameworks for consistency
        - Promotes cross-departmental projects and initiatives
        - Facilitates workshops and forums for knowledge-sharing

2.3. Industrial Gaps

The advent of the new industrial era and accelerating globalisation leads industries to face unexpected changes and challenges, with the current challenges being in the domain of data-driven innovation and developments. These new challenges posed by ever-growing system scales and complexities, inadequate data management, and the exploitation of existing knowledge have jeopardised the entire system's growth and sustainability and, hence, need addressing at the earliest [35–37].

In the era of new start-ups redefining how businesses work, especially in the digital arena, the changes and developments brought about by technology and Industry 4.0 have combined to generate both possibilities and significant hazards for existing industries. The most valuable corporations in the world today are technology businesses. Simultaneously, an overwhelming number of conventional global sectors are susceptible to "Martec's rule", as shown in Figure 2 below, which describes an ever-widening gap between organisational transformation and exponential technological progress. It explains that technology changes exponentially (fast), yet organisations change logarithmically (slow). Technological advancement occurs at a breakneck pace, yet industries move at a snail's pace when it comes to change, adaptation, and evolution. This highlights the difficulties industries encounter in keeping up with new technology as the gap between organisational adaptability and technological capabilities grows over time [38].

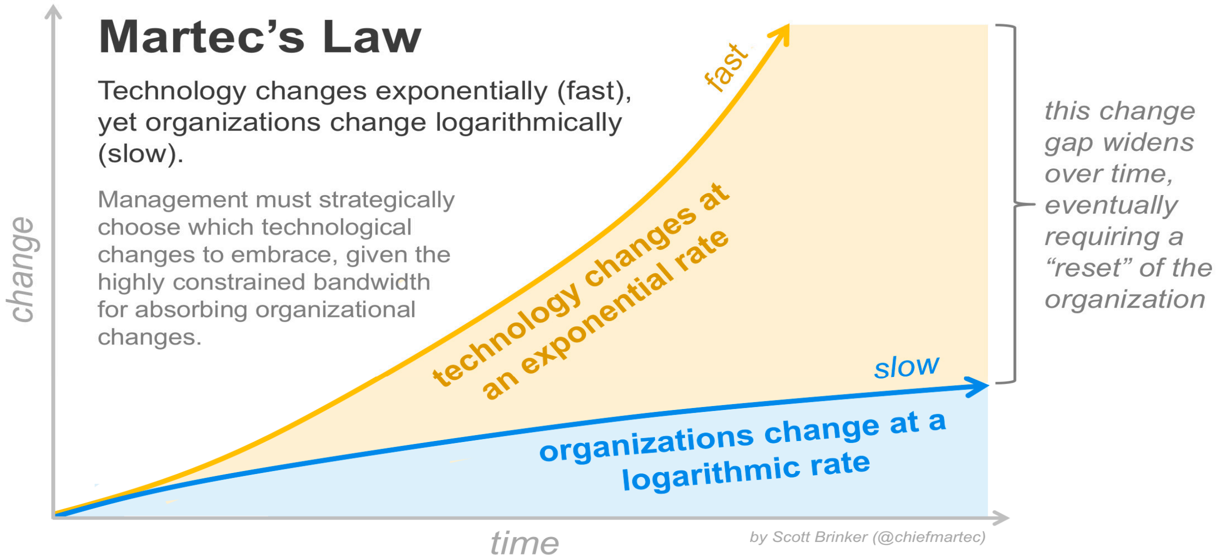

Figure 2. Martec's law [39].

Gap and Why Address It?

Industries in Europe are experiencing difficulty in adapting to the digital landscape. Currently, only 58% of the majority of small and medium-sized businesses (SMEs) in the European Union (EU) have attained a basic degree of digitisation. The number is well below the EU's 2030 aim of 90%. With 91% of them already achieving a basic level of digital services, large businesses with 250 or more employees perform far better. Using at least four of the twelve digital technologies—such as artificial intelligence (AI), cloud computing, the Internet of Things (IoT), social media, and customer relationship management—is considered a basic level of digital services.

The phrase "gap" refers to the difference in level between "where we are" (the current state) and "where we want to be" (the target state). A "maturity level assessment", "requirements analysis", or "need-gap analysis" are all terms used to describe a gap analysis. This process entails determining the present and future states, characterising the gap, and finally bridging the gap between these states, resulting in the industry's development and enhancement [40–42].

Exploring industrial gaps is critical for every industry's advancement to meet current industry standards and client expectations. A gap analysis investigates the effectiveness of a company's information systems, software, or data-driven applications to determine if improvement efforts are being met and, if not, what steps should be made to ensure that they are [43,44].

Although the manufacturing industry has accelerated its adoption of data science and data-driven decision-making in recent years, it continues to face challenges in several areas due to its slow recovery from the recession, slow transformation, and, most importantly, delayed digital technology adoption [45]. It creates several holes in the manufacturing business, particularly in the data sector, some of which are covered in the sections below.

3. Study Design

This study is a narrative literature review. It provides a general overview of industrial data science in the manufacturing sector and highlights the trends and implications of implementing Industry 4.0 and IoT in various industries. This qualitative research seeks to comprehend phenomena by accumulating and interpreting non-numerical data to explore industrial data science and foster a greater understanding of patterns, gaps, and challenges in the existing literature. It gives a direction for advanced and informed decision-making and focuses on the development and research in the field. This narrative review presents a qualitative approach to synthesising existing literature reviews in industrial data science by summarising, analysing, and interpreting existing knowledge, methodologies, and patterns related to industrial data science. The research process mind map is displayed below in Figure 3. This structured research process provides an approach for other researchers if they want to reproduce this study in a similar fashion and can adapt it to their needs for qualitative research.

Figure 3. Study design process.

The initial step was planning. It helped to define the topic, objectives, and scope, which include themes like machine learning applications, optimisation techniques in manufacturing processes, data management, and the role of IoT in data analytics. It also clarified the research question, which indulges the themes of industrial data science, its practices and validation, and the search and selection of the relevant literature, whereby a comprehensive search for appropriate academic manuscripts, such as research papers, articles, conference proceedings and industry reports, white papers, and blogs was conducted. A diverse set of resources providing in-depth insight into industrial data science was gathered and skimmed through for final selection. Databases such as IEEE Xplore, Scopus, and Google Scholar were exploited. Even though the selection process for this narrative review was flexible, it was thorough and kept the industry's view in mind by using the knowledge gained from their white papers and blogs. The critical criteria were considered, such as methodology, findings, and implications to industry practices.

The next step was conducting the research. This included analysing and structuring the literature, which allowed for identifying the literature's trends, gaps, and contradictions. The findings were synthesised to integrate them and create a cohesive narrative to enlighten the research outcome. The research discussed concepts related to data pipelines, analytical capabilities, the impact of data-driven decision-making on manufacturing industries, and the challenges it raises.

The final step was reporting. It included discussion, conclusion, implications, recommendations, limitations, and, most importantly, future research in industrial data science in manufacturing industries. These insights included the best practices for data-driven technology implementation, organisational challenges, and areas that require further exploration.

4. Result and Analysis

4.1. Maturity Level Indexes of the Heavy Machinery Manufacturing Industry

The term "maturity level" refers to a well-defined evolutionary plateau on the way to a fully developed research and innovation process. The word "maturity" refers to the industry's readiness to implement a data strategy [33]. Each maturity level specifies specific process characteristics and provides another layer of security, with higher maturity levels specifying sophisticated features for further process development [46,47]. Each maturity level marks a step forward in pursuing a mature process, establishing a set of objectives that, if achieved, lift an organisation to the next degree of maturity in the process. It also outlines the path taken by a process as it evolves from an immature and ad hoc stage to a mature state. These mostly pertain to the four to six maturity stages that illustrate the progression from novice to reproducible, precise, quantitative, and optimal levels [43,44,48].

Figure 4 illustrates three maturity indices for reaching Industry 4.0 maturity: data capabilities maturity, industrial data science maturity model, and industry data science maturity model. In today's business environment, each maturity level is required to evaluate the industry's present state, which is strongly interconnected. The industry must address and attain three maturity levels to become entirely data-driven [49,50].

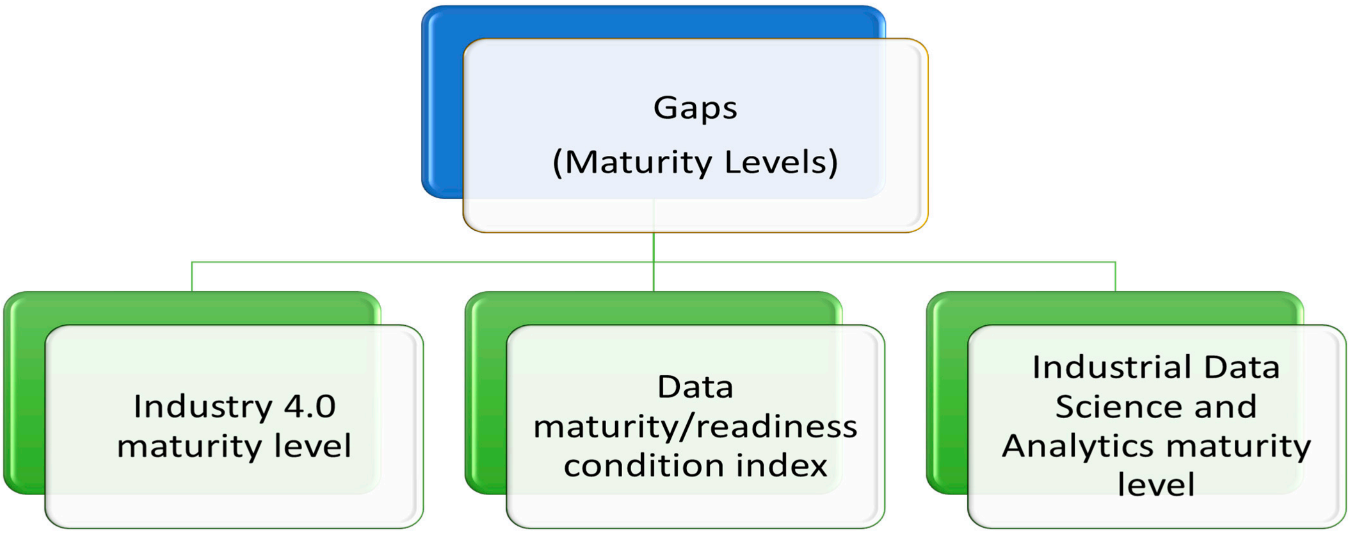

Figure 4. Maturity levels of the manufacturing industry.

A comprehensive description of each of these three types is provided in the following subsections.

4.1.1. Industry 4.0 Maturity Level

Manufacturing sectors must incorporate industrial data science approaches to stay relevant and resilient in the international markets [51]. The industry's ever-changing needs and custom-made goods influence manufacturing and planning processes, forcing organisations to adjust their business strategy to Industry 4.0 rules to stay profitable [52,53]. Businesses that develop comprehensive strategies and assess their current standing in this roadmap may be far more successful across multiple dimensions, including technological development, adaptation and investments, economic performance, societal and environmental impact, and talent pool [52,54,55].

To be successful, each company must analyse its Industry 4.0 maturity levels and take a tiered approach to the Industry 4.0 strategy's fundamental roadmap. A project's maturity level and the maturity model for Industry 4.0 are intertwined in the maturity of the business itself [56–58]:

1. What does the company intend to accomplish—foresight?
2. Evaluate the present conditions—self-evaluation.
3. Where do they wish to go—the endpoint?
4. Missing links must be bridged to get there—the gaps (methodological analysis in Industry 4.0).
5. What strategy to pursue, taking into account procedures, technology, resources, and execution—strategic plan.
6. Evaluation and refinement.

When examining actual machine-equipment operations, production system perspectives, and data, these steps coincide with the good old Data, Information, Knowledge, Wisdom (DIKW) model as follows and must be tackled stage-by-stage [56]

1. Stage 1: Observation of the situation (through the data).
2. Stage 2: Recognise what is happening and why it is happening (through data analysis and knowledge gained from it).
3. Stage 3: Forecasting what will happen (based on previously identified patterns, knowledge, and AI/ML).
4. Stage 4: In the self-optimising Industry 4.0 system, this is the final stage toward total autonomy for self-sufficient equipment. Even for the most advanced data-driven industries today, this vision of Industry 4.0 is far off.

4.1.2. Data Readiness Condition Index

Data is seen as an ancillary asset in every stage of manufacturing, and the advent of IoT devices has significantly reduced the economic burden of data acquisition. On its own, this data deluge is not beneficial. It creates additional value when transformed into actionable information [59]. The data science paradigm has emerged as one of the most promising technologies and research topics in the Industry 4.0 era. The contributing element is the rise in accessible data, which is expected to reach 175 ZB by 2025, as illustrated in Figure 5 below [60]. Industries are becoming increasingly aware of the continuous changes brought on by the massive flood of high-dimensional data, and they are aiming to ride out the storm by investing in their growth [61].

Assessing the industry's data maturity level before starting any data science-based project is a good idea to get the most out of it. However, if it is not completed before or during the proposal stage, it becomes the first step in the project's development [62].

Sensor systems, manufacturing heavy equipment, fault diagnosis systems, Asset Management Systems (AMSs), Information and Communication Technologies (ICTs), various data warehouses, industrial databases, and many other sources generate massive amounts of data at a phenomenal rate in the manufacturing sector [61,63]. The continually growing collection of data science resources boosts the market need for different data science approaches such as data mining, statistical data analysis, analytics, and ML procedures. The type of data determines its applicability. Industry professionals and academics recognise the possibility of automatically extracting insights from this data [64,65].

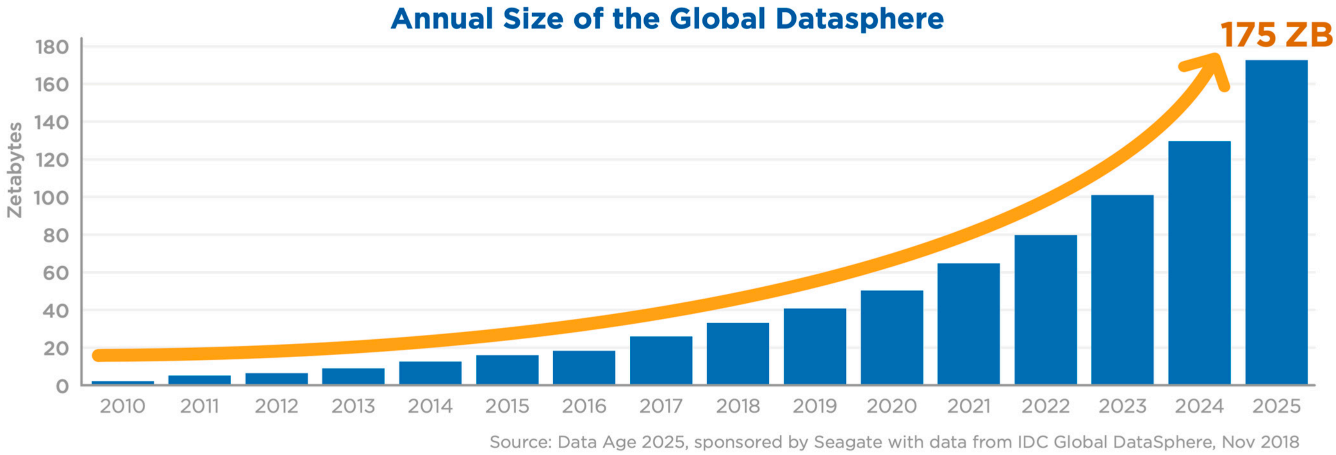

Figure 5. Expected annual growth in global datasphere [60].

Furthermore, it is no secret that data preparedness, let alone the capacity to deal with data and overall data management in the manufacturing industry, should be significantly improved. With existing industry standards, data warehousing, comprehending data, analysing, strategising, security, and collecting data from monolithic systems are all challenges. For example, a lot of data are collected in manufacturing but never make it to data lakes or other systems or repositories where they could be utilised for pattern recognition, ML, DL, or image processing [66]. As a result, it is critical to evaluate the industry's "Data Readiness Condition (Datacon) levels". The Datacon index ranks industries from 0 to 5 in terms of data preparedness [56,60]. Figure 6 below illustrates the various maturity levels the manufacturing industry must achieve to fully embrace Industry 4.0. It highlights the interconnectedness of data capabilities, industrial data science maturity, and overall industry readiness for a data-driven approach.

Another notable approach is the Boston Consulting Group's 'Data Capabilities Model', built on seven data capabilities with the example of specific challenges to be handled, as seen in Figure 7 below. The seven data capabilities are as follows: vision, use cases, data analytics setup, data governance, data infrastructure, data ecosystem, and change management. This index establishes the benchmarks by which industries may evaluate their data capabilities and maturity compared to their rivals. This insight may help companies become more data-driven and improve the design of their digital initiatives [62].

As a result, each business falls into one of the five data maturity levels depicted in Figure 8.

1. Lagging—The organisation is presently encountering a lack of progress or inadequate advancements in the seven data capabilities. This stagnation signifies an inability to confront current difficulties or enact enhancements, leading to a situation where data-related tasks are not performed efficiently.
2. Developing—The organisation acknowledges the significance of its data capabilities and implements specific measures to address issues in this domain. This entails pinpointing particular deficiencies, distributing resources to enhance those competencies, and proactively pursuing solutions to facilitate more efficient data administration and utilisation.
3. Mainstream—The organisation has attained a fundamental level of data proficiency, enabling it to sustain competitiveness with its counterparts. This average performance signifies the successful implementation of fundamental data principles, although there remains potential for growth and enhancement to boost its effectiveness further.

Level	Description	Characteristics	Actionable Steps	Future Trends/Key Metrics
Level 1: Critical	The industry is in disarray, with an uncoordinated ecosystem and no clear vision for third-platform technologies.	1.Lack of leadership within companies. 2.An impending data tsunami exists, with no survival plans in place. 3.Examples: Organizations with significant data challenges and legacy systems.	1.Establish clear leadership and develop survival plans. 2.Assess data strategies and priorities.	Define KPIs for data strategy effectiveness.
Level 2: Needful	The industry is in catch-up mode and struggling to establish a coherent vision.	1.Limited initiatives focused on third-platform technologies, often driven by specific lines of business (LOB). 2.Most generated data is never monetised.	1.Establish cross-functional teams to explore data monetisation strategies 2.Build awareness of benefits among leadership.	Data monetisation rates and project success rates.
Level 3: Average	The industry operates at an average maturity level with a decent ecosystem.	1.Similar visions across companies, with some C-level commitment. 2.Corporate-owned initiatives are in place, often excelling in one area of third-platform technology.	1.Create strategic partnerships to enhance innovation and share best practices.	Assess employee engagement and innovation outcomes.
Level 4: Advanced	The industry is becoming a leader in certain areas of digital transformation.	1.A vibrant ecosystem with several data initiatives owned at the highest management levels. 2.Well-organised data management practices.	1.Invest in employee training programs on data management and digital transformation strategies.	Stay informed about emerging technologies (AI, machine learning).
Level 5: Optimised	The industry is at the forefront of data initiatives, exhibiting a mature data culture.	1.Suppliers are well-aligned and healthy. 2.A data-driven culture exists from the top down, with a skilled workforce executing advanced initiatives. 3.Strong vision for data utilisation.	1.Conduct regular self-assessments against industry benchmarks and adjust strategies.	Continued focus on leveraging data for maximum value extraction.

Figure 6. Industrial data readiness condition index.

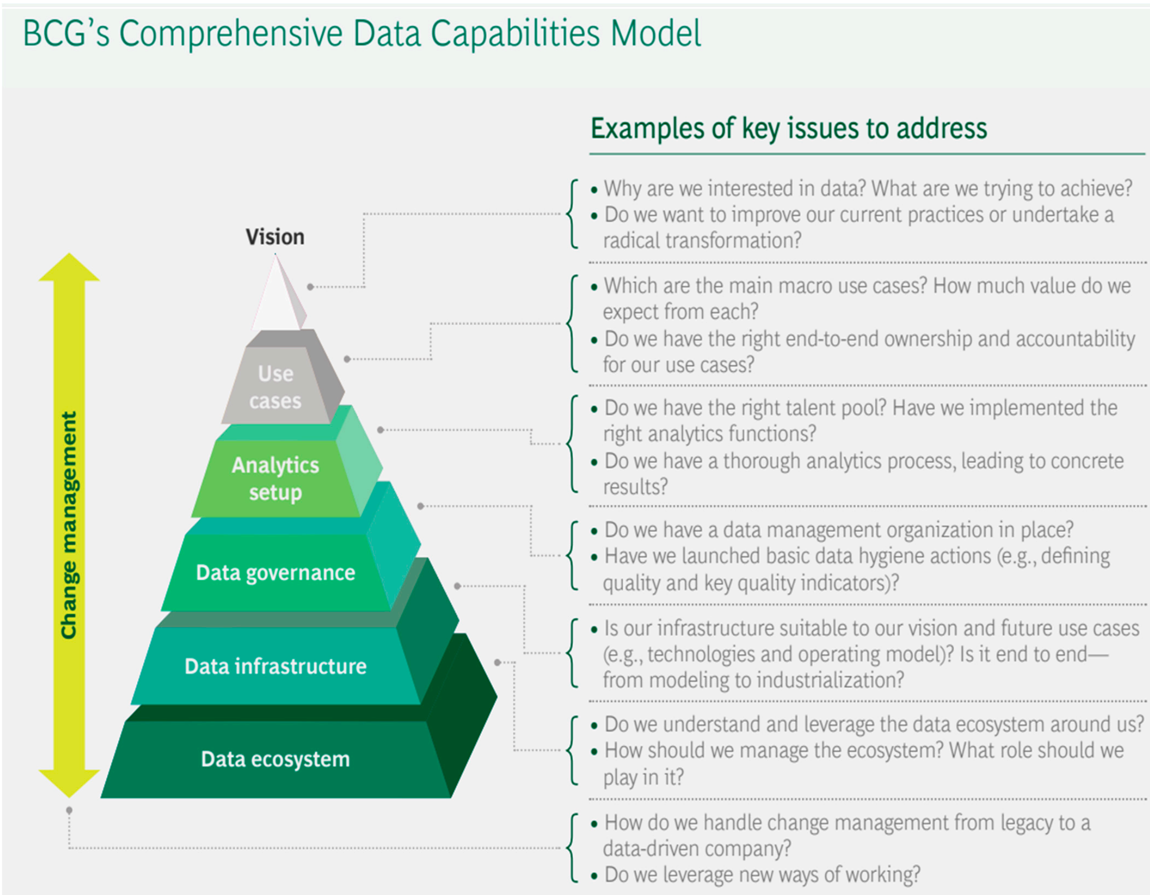

Figure 7. Data maturity or capability assessment model [62].

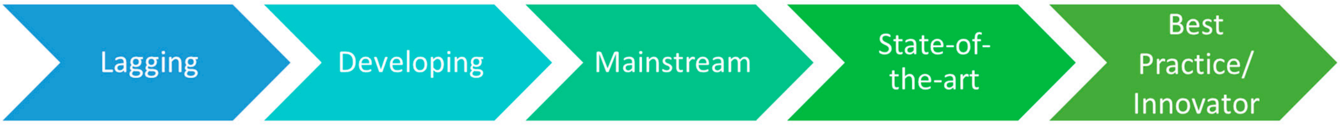

Figure 8. Five-stage data readiness spectrum.

4. State of the art—The organisation demonstrates proficiency in multiple aspects of the seven data capabilities, highlighting sophisticated methodologies and inventive practices. Nonetheless, although it may excel in some areas, it may encounter difficulties in consistently controlling and utilising these talents universally, resulting in potential inefficiencies or overlooked possibilities.
5. Best practice—The organisation distinguishes itself as a leader in data capabilities, exhibiting advanced and efficient practices across all seven domains. It not only upholds these elevated standards but also persistently endeavours to innovate and enhance them, routinely revising strategies, tools, and procedures to accommodate emerging issues and possibilities in the data ecosystem [62].

4.1.3. Industrial Data Science and Analytics Maturity Levels

Several businesses incorporate data analytic tactics into decision-making and action planning to compete in today's progressive and disruptive market. IDS and data analytics are assisting and directing industries in shifting from static to dynamic data-driven decision-makers, allowing them to become more inventive and competitive. The available data determines the industry's status, which enables them to make decisions and take steps to enhance current operations. This decision-making outcome goes from operational to strategic and ultimately to transformative [66,67].

According to research conducted by MIT and IBM, companies with a high level of data analytics had an 8% rise in sales growth, a 24% increase in operational revenue, and a 58% sales increase per person [43]. To progress from a pre-stage or stage zero of no analytics, when the sector has no data analytics procedures at all, save for gathering data for later use or minimum usage and application. Industries follow a five-step trajectory for their analytical development [68–72].

- Descriptive analytics is the process of gathering and visualising historical data to understand past events. It generally involves key performance indicators (KPIs) and metrics. These are often translated into reports, Excel files, and dashboards that only provide the bare minimum of information necessary for customers, rules, and compliance. Otherwise, it creates many reports and data outputs that take a long time to analyse and are frequently neglected.
- Explanatory analytics focuses on deciphering and explaining the why behind the descriptive analytics outcomes. It investigates the factors and offers insight into their relationships and knowledge of why certain things happen and others do not. Techniques like correlations and covariances may be used.
- Exploratory analytics enables data investigation, mainly in its raw form, to describe problems and data gaps and aid in determining the data's quality. To comprehend the narrative behind the data, it uses a combination of data mining, quantitative, and statistical tools to look for patterns, outliers, and abnormalities in the data.
- Diagnostic analytics is a hybrid of exploratory and explanatory analytics. It may be used to find correlations, trends, and other information in existing data and explain why something happened. Even if the beginning purpose and rationale for the application are different, excellent diagnostic models may turn out to be strong predictive models since comprehending the data, correlation, and cause is the stepping stone to prediction.

- Predictive analytics is the process of predicting or forecasting what will happen in the future by analysing large amounts of data using ML technologies. It can forecast future circumstances, devise methods for making better judgments, and have a greater influence. It necessitates a broad grasp of technology, commercial savvy, and more data control.
- Prescriptive analytics emphasises how to make it happen by giving industries the tools to optimise their data, processes, techniques, or even the most basic procedures for decision support and insight generation. It aids in creating a path leading to a specific objective to attain the intended result.

Figure 9 below depicts the five levels of industrial data science and analytics maturity.

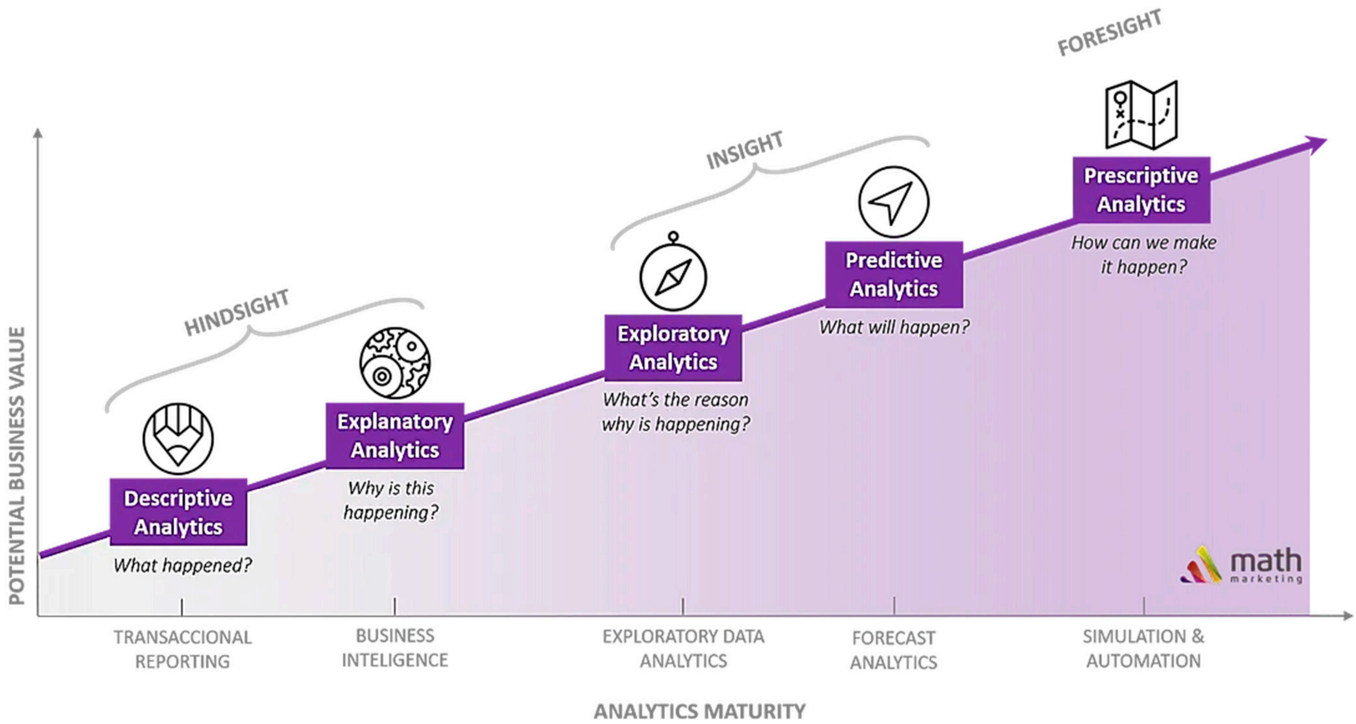

Figure 9. Stages of data analytics maturity [71].

4.2. Industrial Challenges

With the dawn of the new industrial era and the acceleration of globalisation, industries are confronted with unforeseen shifts and problems. The most pressing of which is data-driven innovation and development. These new challenges, which are posed by ever-increasing system scales and complexity, poor data management, and the utilisation of existing information, have jeopardised the entire system's growth and sustainability and must be addressed as soon as possible [73–75].

Industries may encounter a variety of obstacles in many regions. Most of the time, these difficulties are complicated and do not have simple solutions. As illustrated in Figure 10, this article outlines four significant issues that businesses face today in the industrial data science arena.

4.2.1. Using Data Science or Predictive Analysis/Maintenance on Available Industrial Data

Industries are frequently flooded with data gathered from various sources, yet they lack the knowledge to enhance operations and maintain equipment. This condition is known as DRIP—Data Rich but Information Poor. Industries often have abundant data, yet a deficit in insights can hinder effective decision-making. As a result, it is better to gather data for a specified goal to get the most information out of it. Primarily, data evaluation is important to analyse the current data landscape. Pinpointing areas where data collection may be excessive yet unproductive allows for informed adjustments in strategy going forward. The next step is identifying insights by determining the critical data points that foster understanding and inform action. Ensuring data clarity for users to enable them to make informed decisions minimises confusion and oversight. Thirdly, data synthesis is done by compiling essential information regarding data sources, interpretations, patterns, implications for findings, and any contextual factors that may anchor understanding. Finally, one should pinpoint relevant insights, formulate action plans, and establish timelines for executing decisions. This fosters an adaptable culture that utilises insights effectively to maximise opportunities as they arise [75].

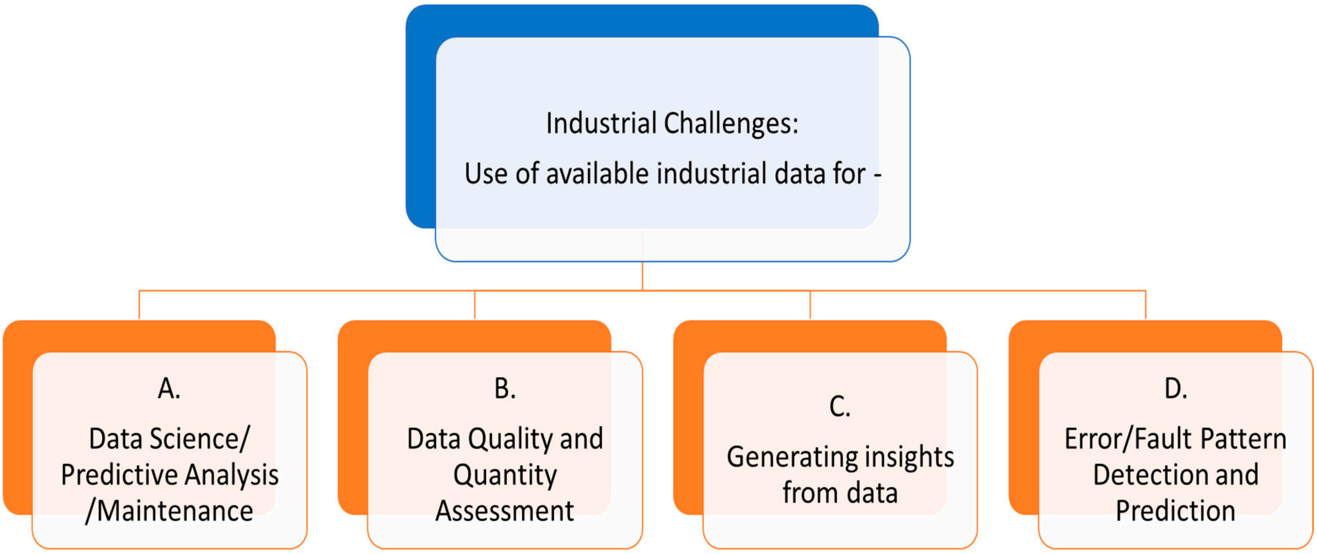

Figure 10. Industrial challenges faced in the industrial data science domain.

Intelligent sensors, networking, and storage have improved current data collection methods, resulting in an ecosystem in which large amounts of data in a variety of formats, including time-series data from industrial processes, assets, devices, suppliers, customers, and other sources, are accumulated [75,76].

For every data-related project, discovering datasets that may be used for data analytics or predictive analytics for future use in maintaining and improving a valuable industrial asset becomes extremely difficult [77,78]. On the other hand, these data sources may provide significant insights into the status and operations of virtually all of the industry's major equipment and critical processes. It may even be useful for prediction, but this is highly dependent on how and for what purpose the data is collected and whether or not, in cases where multiple data sources are involved, the data can be synchronised to achieve the best results [79–82].

Industries require data scientists to enable their quest into the data world, uncover knowledge, turn it into insights, and present it using data storytelling to industrial stakeholders to make meaningful decisions for greater profit.

According to the Financial Times, many industries underutilise data scientists because they lack the raw resources needed to produce outcomes. According to a Stack Overflow study, 13.2% of data scientists, second only to ML professionals, want to leave the industry because of their obstacles, such as a lack of raw data and resources [83–85].

According to Goldbloom, those who work in this field face several challenges. One of the main reasons is that their employers cannot offer them the essential raw materials to produce results; also, some complain about a lack of precise questions to respond to.

While companies may see the opportunity, they often lack the understanding required to maximise the value of their data assets. It also highlights the technical knowledge gap between non-specialist managers, data scientists, and ML experts [83,84].

Furthermore, data science poses several issues due to its ambiguous limits, the subject's multi-dimensionality, and the fact that it has seen quick and unexpected improvements in recent years. New definitions, methods, software, subject areas, and other factors are increasing the complexity of data science, which poses several issues [86]. It is vital to be aware of potential stumbling blocks to learn how to overcome them. The following are only a few of the issues that data science presents [87]:

- It is inherently multi-disciplinary
- It has no well-defined scope
- It has much overlapping terminology
- It has no stipulated boundaries
- Used with technical or popular connotations
- Used with limited or broad meanings
- Slightly altering in time
- Data quality and problems linked with it
- Data unavailability and inaccessibility is another key issue
- Sudden and quick technical advancements as a result of the following:
    - Increased computing capabilities
    - Increased data availability with substandard quality
    - Incorporation of new areas, tools, and more without proper education
- Due to the aforementioned factors, mastering data science is exceedingly arduous

A total of 10,153 people participated in Kaggle's 2017 State of Data Science and ML poll. It consisted of five main categories, and the data professional's responses reveal the challenges they face [88,89].

4.2.2. Data Quality and Quantity

The word "data quality" refers to high-quality data that are (1) appropriate for their intended application and (2) closely related to the hypothesis being evaluated [90]. The increasing volume of data accessible today and varying degrees of quality have made it increasingly difficult to analyse and use data critical to the organisation efficiently [91]. This definition emphasises that, in addition to the ability to use and access data, data quality is greatly influenced by the context of data use and synchrony with user expectations. As a result, the data quality review and enhancement process should include data users and other stakeholders involved in data acquisition, processing, and analysis [90–93].

Additionally, businesses that create "Big Data" and associated analytics have garnered considerable interest in extracting information, knowledge, and wisdom. Troves of high-dimensional, streaming, and unstructured data are gathered and processed in the industry to aid decision-making. This data may be combined to provide insights, recommendations, and actions. Data can, however, be structured, semi-structured, unstructured, untidy, disorganised, and missed [94,95].

SMEs may improve their capacity to improve productivity, integrate fact-based insights, and make more successful choices by implementing a solid data analytics strategy. However, such tactics are only as good as the data utilised to drive them [96]. Furthermore, creating a user-friendly platform that supports collaboration and individualised problem-solving for SMEs is essential to conducting effective data analytics and skill development. A robust technical infrastructure is necessary for data analysis and learning. Accessibility ensures SMEs can easily integrate and utilise analytical results, while security protects personal and company data throughout the analysis process [97].

A typical data analysis dilemma is that large amounts of industrial data are frequently created and gathered without regard to a clear goal or target for the data collection and subsequent use. Different datasets are generated in the industry from multiple departments, assets, or procedures, typically without data warehousing. Organisations' inability to offer high-quality data has resulted in many issues, including erroneous judgments based on inaccurate data, excessive operating expenses, and a lack of customer satisfaction [93].

To better comprehend the industry and the data, it is necessary to examine numerous data streams [98,99]. However, when the data itself is not well known or remains substantially unexplored, there is no such thing as a "one size fits all" approach to data quality evaluation applied to these datasets and databases [100,101]. Furthermore, such raw data might be inconsistent or missing [102].

The universal, two-layer criterion for assessing data quality in diverse datasets to determine their feasibility for further processing is shown in Figure 11. It aids in determining whether the relevant data sources are of sufficient quality for the objective [93].

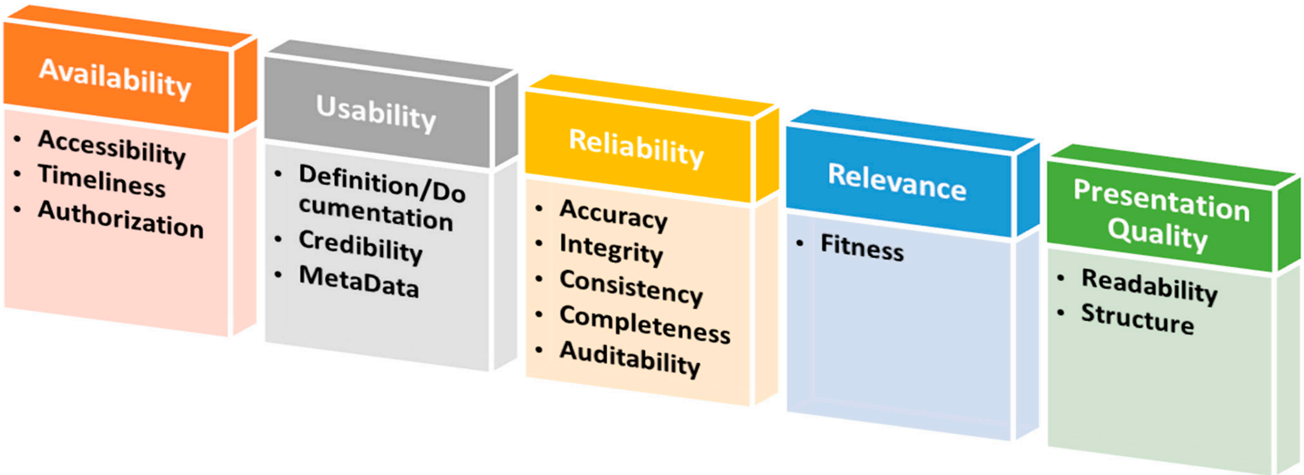

Figure 11. A universal, two-layer standard for assessment of data quality [93].

4.2.3. Using Data to Generate Insights

The fundamental goal of data gathering is to analyse information to produce insights for companies that will help them decrease asset downtime and improve efficiency, revenue growth, and overall performance. Basic observations and information that an industry acquires through analysing its dataset are known as data insights. Data are fed into a system, and insights are produced. That assists them in making educated decisions and mitigates the risks involved with trial-and-error testing [103,104]. The following are some of the elements that influence actionable insights:

- The availability of high-quality data
- The parameters, characteristics, or features of a dataset
- The specific problem to be addressed
- What information should be used to gain a competitive advantage?

If any aspect of this is not addressed, it proves to be a challenge for the industry. However, not all insights generated through industrial data analysis are actionable, which can be challenging. Some insights can be used for informational or remedial purposes and then translated into actionable tasks when processed [105,106]. As illustrated in Figure 12, data insights, in whatever form, may substantially benefit the industry's advancement.

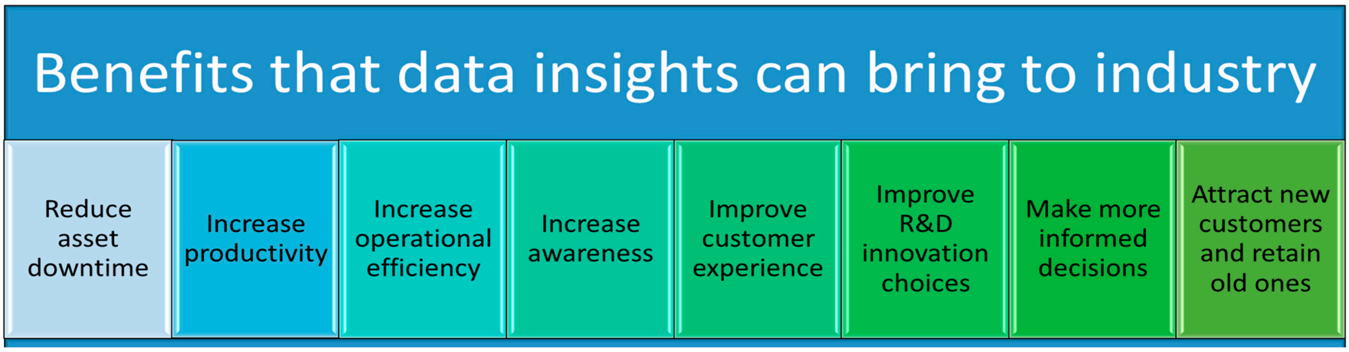

Figure 12. Data insight benefits.

4.2.4. Error Pattern Recognition and Prediction

The tremendous expansion of digitalisation leads to the global data upswing. Globally created, recorded, copied, and consumed data were predicted to reach 64.2 zettabytes by 2020. They will surpass 181 zettabytes in 2025, and the manufacturing industry contributes a major portion of this data [107]. However, most of this is raw data. This data may advance knowledge and insights into various vital elements, ushering in the big data age. The enormous amount (volume), rate of collection (velocity), and variety of data are the core qualities that distinguish Big Data; nevertheless, additional features are constantly being created, with the value being the most significant [108].

One of the essential needs for manufacturing industries is turnaround time, largely dependent on the industrial equipment's short downtime, efficiency, and performance. A direct correlation exists between reduced errors and increased productivity, efficiency, turnover, and shorter downtime [109,110]. As a result, a lot hinges on decreasing errors and detecting them early enough to take remedial action whenever warranted. As a result, it is critical to identify errors that might turn into faults if they go undiscovered and pose risks of failure during operation [111].

In addition, establishing predictions requires a thorough understanding of pattern recognition [112]. Pattern detection and recognition are crucial for establishing whether an error is a one-time occurrence or a steady deterioration of the condition before the failure. Pattern recognition of error-related data aids in the identification of trends and regularities across time [113]. Furthermore, using ML and DL techniques may aid in detecting and predicting erroneous or anomalous patterns.

Therefore, it is critical to determine which industrial data sources are appropriate for deciphering error data. It is necessary to review and analyse various datasets and databases that the industry may have to get better results for predictions [114,115].

5. Discussion

Industries, especially in the manufacturing sector, are transforming significantly as they increasingly embrace data-driven approaches. They meticulously collect and store data from various sources and assets to derive valuable insights for making informed decisions [72]. By leveraging industrial data science, these data can be utilised for a multitude of purposes, such as predictive maintenance to anticipate and address equipment failures, implementing just-in-time spare part procurement based on error and fault predictions, enabling effective management decisions, and driving significant cost savings. Moreover, this knowledge can be instrumental in guiding system upgrades to enhance overall operational efficiency and performance.

This article comprehensively explores industrial data science and its application in data-driven manufacturing industries. It also delves into the intricate process of developing a data-driven culture within industries and the various levels of maturity associated with this transition. The surge in industrial data science adoption can be primarily linked to the fourth industrial revolution, Industry 4.0, and the disruptive technologies accompanying it, such as the Industrial Internet of Things (IIoT), cloud computing, sensor technology, and automation for autonomous machines. Table 1 below outlines key factors related to gaining competitive advantage through data insights in various industries, highlighting the importance of error pattern recognition and effective data management. It emphasises the transformative role of data-driven approaches in enhancing performance, reducing downtime, and optimising decision-making.

Table 1. Key insights into competitive advantage and data-driven transformation in the industry.

Section	Content
Competitive Advantage	What information should be used to gain a competitive advantage? If any aspect of this is not addressed, it will prove to be a challenge for the industry. Some insights can be informational or remedial, translated into tasks later.
Data Insights Benefits	Data insights benefit the industry's advancement significantly.
Error Pattern Recognition	The expansion of digitalisation led to the creation of global data, which was 64.2 zettabytes by 2020 and surpassing 181 zettabytes by 2025. Most of it is raw data, essential for pattern detection and improvement in manufacturing.
Turnaround Time	A need for short near-zero equipment downtime to reach 100% performance, with a correlation between reduced errors and increased efficiency. Addressing errors early is crucial for manufacturing.
Pattern Recognition	Identifying whether errors are isolated incidents or signs of deterioration. ML and DL techniques help identify anomalies.
Data Sources	Determining appropriate industrial data sources for error data analysis is necessary for better predictions.
Industry Transformation	Industries are adopting data-driven approaches to derive insights for informed decision-making and predictive failure. This will reduce maintenance and cost savings by avoiding unwanted, costly spare inventory.
Data Management Challenges	Emphasising data preparedness and assessing data capabilities through benchmarks like DATACON levels and the Data Capabilities Model.
Adoption of Data Science	Linked to the fourth industrial revolution and technologies such as IIoT and automation.
Analytical Techniques	Techniques include data mining, analytics, AI, ML, DL, big data processing, and advanced mathematical modelling.

Furthermore, this study sheds light on the challenges associated with data management in the manufacturing industry, emphasising the importance of data preparedness and the capacity to deal with data. It discusses the concept of the "Data Readiness Condition (DATACON) levels" and the "Data Capabilities Model" as benchmarks for evaluating data capabilities and maturity in comparison to industry rivals.

The exponential growth in global data collection calls for sophisticated analysis techniques encompassed within the industrial data science framework. These techniques span various technical disciplines, including data mining, data analysis, data analytics, statistical analysis, AI, ML, DL, neurocomputing, big data processing, pattern recognition, and advanced mathematical modelling. From a business standpoint, industrial data science also encompasses business intelligence (BI), business analytics, business development strategies, storytelling using data, and data visualisation techniques to communicate insights to stakeholders and decision-makers effectively.

The impact of the study is evident in its insights on the five-stage data readiness spectrum, industrial data science and analytics maturity levels, and the correlation between data analytics and business performance. It provides a roadmap for industries to progress through the five levels of analytical development and highlights the benefits of incorporating data analytics into decision-making processes.

Overall, the study's relevance lies in its contribution to guiding manufacturing industries towards becoming more data-driven, competitive, and innovative through effective data management and utilisation. It underscores the transformative potential of data analytics in driving sales growth, operational revenue, and overall business performance. Table 2 below presents the key aspects and findings:

Table 2. Key aspects and findings.

Key Aspect	Findings
Challenges in Data Science	Lack of clear boundaries, evolving definitions, quality issues, data unavailability, and rapid tech advancements.
Data Quality	Critical for efficient analysis; influenced by context and user expectations.
Data Quantity Issues	Industrial data often collected without clear objectives, leading to erroneous judgments and high costs.
Insights Generation	Data must be high-quality and relevant to produce actionable insights that enhance decision-making.
Error Pattern Recognition	Digitalisation drives data creation, necessitating advanced analytical capabilities to predict errors.

6. Conclusion and Future Directions

This paper states that industrial data science enables enterprises to develop from infancy to the ultimate data innovation leader stage. However, this transformation is complex for the manufacturing industries as several are not yet automated or digitalised. It brings numerous challenges, which are identified and addressed in this article, focusing on data.

This study suggests the four primary areas of challenges are as follows and gives comprehensive information about them:

- Data science—predictive analytics/maintenance
- Data quality and quantity assessment.
- Insights generation from data
- Error/fault pattern detection and prediction

Manufacturing industries must identify and address shortcomings to stay competitive with other domains. It will enable them to enhance their abilities and become more data-driven, aligning with the direction of the global market. Understanding the maturity level gaps in various aspects of the industry is crucial.

This paper outlines three maturity-level gaps that the industry currently faces:

- Industry 4.0
- Data readiness condition index, also known as data maturity
- Industrial data science and analytics maturity

This paper emphasises the need for manufacturing industries to transition from lagging or stagnant data capabilities to innovative, high-tech, and efficient data analytics to become truly data-driven. This paper also highlights the challenges data scientists face within the industry and the benefits of data insights to the manufacturing sector. It discusses the vital role ML and DL's implementation can play in leading innovations in the data domain for pattern recognition and prediction.

The main findings of this paper are as follows:

- Widening gap in digital maturity: This paper highlights the widening gap in digital maturity among companies due to the rapid adoption of the latest technologies. It emphasises the "need gaps" or maturity levels where industries must leverage growing opportunities. These gaps are divided into three maturity levels: Industry 4.0 adoption, data readiness (data maturity), and industrial data science and analytics capabilities. This finding implies that industries must leverage growing opportunities and develop specialised strategies to integrate and use digital technologies effectively.
- Challenges associated with data utilisation: This paper addresses the challenges associated with data collection, storage, and application in industrial settings, emphasising the need for robust industrial data science strategies to overcome these obstacles. This finding implies the necessity of developing comprehensive strategies to harness the potential of industrial data science and analytics for driving industrial advancement in the manufacturing industry.

The implications of the main findings are as follows:

- Strategic roadmap: The widening gap in digital maturity underscores the need for companies to develop specialised strategies to integrate, use digital technologies effectively, and stay competitive in the rapidly evolving digital landscape. This highlights the importance of investing in the necessary infrastructure and expertise to bridge these gaps and fully leverage the potential of Industry 4.0 and data-driven decision-making.
- Competitive advantage: The challenges associated with data utilisation emphasise the need for robust industrial data science strategies to overcome data collection, storage, and application obstacles. Addressing these challenges can lead to improved operational efficiency, cost reduction, and strategic decision-making, ultimately contributing to the competitiveness and profitability of manufacturing industries. Overall, these findings underscore the critical role of digital transformation and data-driven strategies in shaping the future of manufacturing industries.

The application of data-driven manufacturing does present several limitations, as acknowledged below:

- Scope of data: When implementing data-driven strategies, not all probable data sources or types may be considered, potentially limiting the insights' comprehensiveness. Specific industries might have unique data challenges that are not fully addressed.
- Rapid technological changes: The pace of advancement in technology, such as AI, ML and IIoT, can quickly outdate certain findings, making it difficult to maintain the relevance of recommendations and implications.
- Sector variability: Different manufacturing sectors might face distinct challenges and benefits associated with data-driven approaches, which generalised studies might not completely comprehend.
- Data privacy and security concerns: This study does not delve into the data privacy and security implications as industries heavily rely on data collection, which is a crucial collection.
- Cultural resistance: Significant internal resistance to transformation to a data-driven culture can exist. The organisational behaviour patterns and factors that hinder data adaptation may not be sufficiently addressed.
- Resource allocation: The importance of resource allocation and investment in technology infrastructure, which are critical for successfully implementing data-driven strategies, may be pretermit.

- Outcome measurements: Clear metrics for measuring the success of data-driven initiatives should be established; the study might not address this comprehensively in this paper.

Innovation policies are essential for promoting creativity and technical progress in all organisations and industries. To enhance their efficacy, it is crucial to adopt best practices that foster a conducive atmosphere for innovation. Some suggestions for best practices for innovation policies are as follows:

- R&D investments: Foster increased R&D investments to explore emerging and disruptive technologies and their applications within specific manufacturing contexts. Government grants and incentives can encourage private-sector investments.
- Skill development programs: Promoting education and training programs to equip the workforce with the necessary data analysis, ML, and data management skills by collaborating with academic institutions and enhancing relevant curricula.
- Data governance frameworks: A framework for robust data governance policies that ensure data privacy, security, ethical data use, and fostering trust in data initiatives across organisations must be established.
- Collaboration and partnerships: Encouraging collaboration among stakeholders, including private companies, research institutes, and government bodies, to share knowledge, resources and best practices in technology adoption.
- Pilot programs: These programs allow industries to experiment with data-driven strategies on a smaller scale before full implementation, minimising risk and allowing for adjustment based on outcomes.
- Funding and grants: To enable a more widespread transaction across different industry segments by providing grants or funding opportunities for small and medium enterprises (SMEs) to adopt data-driven technologies.
- Creating a culture of innovation: Greater openness towards adopting new technologies can encourage a culture that supports experimentation and acceptance of failure as a learning opportunity.
- Incentivising innovation: Formulate guidelines that incentivise enterprises to effectively incorporate data science into their operations, fostering competitiveness and knowledge dissemination.
- Monitoring and evaluation: Develop a framework for continuous monitoring and evaluation of technology adaptation initiatives to evaluate progress, results, and areas requiring enhancement.
- Public awareness campaigns: Initiate campaigns to enhance awareness of data-driven processes' advantages and transformative capacity, fostering acceptability among the workforce and the broader community.

By overcoming these obstacles and enacting new policies, industries can more effectively facilitate the development, adaptation and dissemination of technologies underpinning industrial data science and enhance their competitiveness in a changing environment.

Furthermore, future research should include more investigation into practical ways of resolving the challenges outlined in the paper. It could consist of creating particular guidelines or best practices for applying data science, assessing data quality and quantity, increasing insights creation, and improving error/fault pattern detection and prediction in the manufacturing industries.

Furthermore, future research should develop and implement Industry 4.0 technologies and techniques, improve data readiness conditions, and increase industrial data science and analytics maturity. It could include conducting case studies, pilot projects, and collaborative initiatives to test and refine novel approaches to resolving these maturity gaps.

Furthermore, future efforts should focus on providing tangible support and resources to manufacturing companies as they migrate to more advanced data capabilities. It could include creating training programs, knowledge-sharing platforms, and consulting services to assist businesses in implementing innovative, cutting-edge, and efficient data analytics techniques.

Additionally, research should focus on the role of ML and DL in driving advances in data analytics for pattern detection and prediction in the industrial sector. Research and development in this field could result in the development of specialised ML and DL solutions that suit the industry's specific demands and difficulties, further advancing its data-driven transformation. This work is already underway, and an example is DL-based implementation in this area describing marine logistics and container crane error detection, diagnosis, and prediction using a custom algorithm that employs Synthetic Minority Oversampling TEchnique (SMOTE) and Long Short-Term Memory (LSTM) [116].

A case study highlighting a manufacturing industry in the container shipping domain is ongoing. The study will showcase the company's progression from descriptive to prescriptive analytics, from the infancy or lagging stage to the market leaders and best practice innovators, to demonstrate the theory presented in this paper to a practical approach. It will demonstrate the practical application of the theory presented in this paper.

Author Contributions: Conceptualisation, A.A.; data curation, A.A.; formal analysis, A.A.; methodology, A.A.; software, A.A.; supervision, L.K. and J.W.; validation, A.A.; visualisation, A.A.; writing—original draft, A.A.; writing—review and editing, A.A., L.K. and J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors gratefully acknowledge their colleagues in the IMaR Research Centre and the Research Office at Munster Technological University, Tralee, and their industry partners for their continued support and permission to publish this research. This work was supported, in part, by Science Foundation Ireland grant 13/RC/2094 and co-funded under the European Regional Development Fund through the Southern and Eastern Regional Operational Programme to LERO, the Science Foundation Ireland Research Centre for Software (www.lero.ie).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the study's design; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. IPS Engineers GMbH. Industrial Data Science—IPS Engineers | Fundiert. Innovativ. Operativ. IPS Engineers GmbH. Available online: https://ips-engineers.com/portfolio/industrial-data-science-en/ (accessed on 23 March 2022).
2. Montgomery, O. 2024 Tech Trends in Manufacturing: Insights to Attract and Retain Software Buyers. Available online: https://www.gartner.com/en/digital-markets/insights/2024-tech-trends-in-manufacturing (accessed on 10 June 2024).
3. Zhang, X.; Ma, L.; Peng, K.; Zhang, C.; Shahid, M.A. A cloud–edge collaboration based quality-related hierarchical fault detection framework for large-scale manufacturing processes. *Expert. Syst. Appl.* **2024**, *256*, 124909. [CrossRef]
4. Salam, A.; Ullah, F.; Amin, F.; Abrar, M. Deep Learning Techniques for Web-Based Attack Detection in Industry 5.0: A Novel Approach. *Technologies* **2023**, *11*, 107. [CrossRef]
5. Zolkin, A.L.; Aygumov, T.G.; Matvievskaya, T.B.; Bityutskiy, A.S.; Dragulenko, V.V. The impact of industry 4.0 on automation and information technology. *AIP Conf. Proc.* **2024**, *3102*, 30010. [CrossRef]
6. Chu, F.; Gailus, S.; Liu, L.; Ni, L. The Future of Port Automation | McKinsey. Travel, Logistics & iInrastructure McKinsey&Company. Available online: https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/the-future-of-automated-ports (accessed on 24 May 2021).
7. GE Intelligent Platforms. The Rise of Industrial Big Data. GE Intelligent Platforms. Available online: https://silo.tips/download/the-rise-of-industrial-big-data-2 (accessed on 23 March 2021).

8. Bourgi, S.; Del Castillo, D.; Freund, G. *Big Data & The Intelligence Economy Canada's Hyper Connected Landscape Information and Communications Technology Council (ICTC)*; Information and Communications Technology Council: Ottawa, ON, Canada, 2015.
9. Baily, M.; Manyika, J. *Is Manufacturing 'Cool' Again?* McKinsey Global Institute: Brussels, Belgium, 2013.
10. Nuur, A. Performance Prediction Model for a Ship-to-Shore Crane: Productiviteit Predictie Model Van Een Schip Tot Wal Kraan. Master's Thesis, Delft University of Technology, Delft, The Netherlands, 2020.
11. Zrnic, N.; Petković, Z.; Bošnjak, S.M. Automation of ship-to-shore container cranes: A review of state-of-the-art. *FME Trans.* **2005**, *33*, 111–121.
12. Horrell, M.; Reynolds, L.; McElhinney, A. Data Science in Heavy Industry and the Internet of Things. *Harv. Data Sci. Rev.* **2020**, *2*, 1–13. [CrossRef]
13. Sluzki, N. 'Data Integrity Issues: Examples, Impact, and Prevention | IBM', ibm.com. Available online: https://www.ibm.com/think/insights/data-integrity-strategy (accessed on 4 January 2024).
14. Manyika, J.; Chui, M.; Miremadi, M.; Bughin, J.; George, K.; Willmott, P.; Dewhurst, M. *Harnessing Automation for a Future That Works*; McKinsey Global Institute: Brussels, Belgium, 2017.
15. Delenclos, F.-X.; Rasmussen, A.; Riedl, J. To Get Smart, Ports Go Digital. Boston Consulting Group (BCG). Available online: https://www.bcg.com/publications/2018/to-get-smart-ports-go-digital (accessed on 20 June 2021).
16. Paraszczak, J. Failure and downtime reporting—A key to improve mine equipment performance. *CIM Bulletin* **2000**, *93*, 73–77. Available online: https://www.researchgate.net/publication/297902012_Failure_and_downtime_reporting_-_A_key_to_improve_mine_equipment_performance (accessed on 27 June 2024).
17. Sun, D.; Hu, J.; Wu, H.; Wu, J.; Yang, J.; Sheng, Q.Z.; Dustdar, S. A Comprehensive Survey on Collaborative Data-access Enablers in the IIoT. *ACM Comput. Surv.* **2023**, *56*, 1–37. [CrossRef]
18. Petzelt, D.; Schallow, J.; Deuse, J. Data integration in digital manufacturing based on application protocols. In Proceedings of the 2010 3rd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2010, Chengdu, China, 9–11 July 2010; Volume 2, pp. 475–479. [CrossRef]
19. Industrial Data Scientist: The New Limb of Industrial Workforce. Available online: https://www.dasca.org/world-of-data-science/article/industrial-data-scientist-the-new-limb-of-industrial-workforce (accessed on 9 September 2024).
20. Mecalux. Industry 4.0 vs. Industry 5.0—Mecalux.com. Industrial Website. Available online: https://www.mecalux.com/blog/industry-4-0-vs-industry-5-0 (accessed on 19 September 2024).
21. Casas, P. *Industrial Data Science—AIT Austrian Institute of Technology*; AIT Austrian Institute of Technology GMBH: Vienna, Austria, 2019.
22. Khine, P.P.; Shun, W.Z.; Khine, P.P.; Shun, W.Z. Big Data for Organizations: A Review. *J. Comput. Commun.* **2017**, *5*, 40–48. [CrossRef]
23. Mourtzis, D.; Vlachou, E.; Milas, N. Industrial Big Data as a result of IoT adoption in Manufacturing. In Proceedings of the 5th CIRP Global Web Conference Research and Innovation for Future Production—ScienceDirect, Patras, Greece, 4–6 October 2016; pp. 290–295. [CrossRef]
24. Deuse, J.; West, N.; Syberg, M. Rediscovering Scientific Management—The Evolution from Industrial Engineering to Industrial Data Science. *IJPME* **2022**, *10*, 1–12. [CrossRef]
25. Espina-Romero, L.; Noroño Sánchez, J.G.; Gutiérrez Hurtado, H.; Dworaczek Conde, H.; Solier Castro, Y.; Cervera Cajo, L.E.; Rio Corredoira, J. Which Industrial Sectors Are Affected by Artificial Intelligence? A Bibliometric Analysis of Trends and Perspectives. *Sustainability* **2023**, *15*, 12176. [CrossRef]
26. TrendMiner. Industrial Data Analytics in the Manufacturing Industry—TrendMiner. Available online: https://www.trendminer.com/big-data-analytics-in-manufacturing-industry/ (accessed on 25 July 2021).
27. Deuse, J.; Schmitt, J. Industrial Data Science: The Benefits of Artificial Intelligence for Production—KAN Kommission Arbeitsschutz und Normung. Available online: https://www.kan.de/en/publications/kanbrief/setting-the-course-for-work-equipment-of-the-future/industrial-data-science-the-benefits-of-artificial-intelligence-for-production/ (accessed on 25 July 2021).
28. Digital, G.E. Part 2: From Data Science to Industrial Data Products | GE Digital. Available online: https://www.ge.com/digital/blog/part-2-data-science-industrial-data-products (accessed on 25 October 2020).
29. Amruthnath, N. Data Science in Manufacturing: An Overview. Available online: https://opendatascience.com/data-science-in-manufacturing-an-overview/ (accessed on 20 September 2024).
30. Durcevic, S. Data Science Applications in Various Industries—DataScienceCentral.com. Available online: https://www.datasciencecentral.com/data-science-applications-in-various-industries/ (accessed on 20 September 2024).
31. NYIT Data Science. Industry-Specific Data Science Applications: Insights and Impact | NYIT Data Science. Available online: https://online.nyit.edu/blog/industry-specific-applications-of-data-science (accessed on 20 September 2024).
32. Chawre, H. Data Science Application Redefining Industries in 2023. Available online: https://www.turing.com/kb/how-is-data-science-application-transforming-industries (accessed on 20 September 2024).

33. Leiras, A.; González-Calderón, C.A.; de Brito, I., Jr.; Villa, S.; Yoshizaki, H.T.Y. *Operations Management for Social Good: 2018 POMS International Conference in Rio*; Springer: Berlin/Heidelberg, Germany, 2019.
34. Hashim, M.A.; Ibrahim, C.K.I.C.; Jaafar, N.A.L.; Kordi, N.E.; Haron, A.T.; Umeokafor, N. Building data driven culture for digital competitiveness in construction industry: A theoretical exploration. *Int. J. Constr. Manag. March* **2024**, 1–13. [CrossRef]
35. Adland, R. Unlocking the Full Value of Maritime Data Science | Windward. In *Sea: The Future' Conference Magazine*; Windward: Tel Aviv, Israel, 2019; p. 3.
36. Pacific Tycoon, Understand the Elements of The Container Shipping Industry | by Pacific Tycoon | Medium. Available online: https://medium.com/@PacificTycoon/understand-the-elements-of-the-container-shipping-industry-a6f5652af907 (accessed on 22 March 2021).
37. United Nations Conference on Trade and Development (UNCTAD). *50 Years of Review of Maritime Transport 1968–2018—Reflecting on the Past, Exploring the Future Transport and Trade Facilitation Series No.11*; UNCTAD: Geneva, Switzerland, 2018.
38. Kupilas, K. Industry 4.0 and Digital Maturity. *Industry 4.0 and Project Management and Engineering*. Available online: https://www.academia.edu/44850284/Industry_4_0_and_Digital_Maturity (accessed on 28 November 2024).
39. Martec's Law: The Greatest Management Challenge of the 21st Century—Chief Marketing Technologist. Available online: https://chiefmartec.com/2016/11/martecs-law-great-management-challenge-21st-century/ (accessed on 5 December 2024).
40. Greenblatt, D. *6 Gaps the Manufacturing Industry Needs to Close to Move Forward*; LinkedIn: Sunnyvale, CA, USA, 2015.
41. Sales, F. What Is a Gap Analysis?—Definition from WhatIs.com. Available online: https://searchcio.techtarget.com/definition/gap-analysis (accessed on 22 July 2021).
42. Osterman, C. *Defining Gaps in Lean Increasing the Ability to Solve Problems in a Production System*; Mälardalen University: Västerås, Sweden, 2020.
43. Mertens, J. Unlock the People Equation | IBM. Available online: https://www.ibm.com/thought-leadership/institute-business-value/report/people-equation (accessed on 20 July 2021).
44. Goldenson, D.; Jarzombek, J.; Rout, T. Measurement and analysis in capability maturity codel: Integration models and software process improvement. *Crosstalk J. Def. Softw. Eng.* **2003**, *6*, 20–24.
45. Gregolinska, E.; Khanam, R.; Lefort, F.; Parthasarathy, P. Industry 4.0: Digital Transformation in Manufacturing | McKinsey. McKinsey and Company. Available online: https://www.mckinsey.com/capabilities/operations/our-insights/capturing-the-true-value-of-industry-four-point-zero (accessed on 4 January 2024).
46. Sanchez, L.; Delgado, A.; Ruiz, F.; Garcia, F.; Piattini, M. Measurement and Maturity of Business Processes. In *Handbook of Research on Business Process Modeling*; IGI Global: Hershey, PA, USA, 2009; pp. 532–556. [CrossRef]
47. Carnegie Mellon University. *CMMI Maturity Levels*. Tutorialspoint. 2019. Available online: https://www.tutorialspoint.com/cmmi/cmmi-maturity-levels.htm (accessed on 20 July 2023).
48. Almeida, J.; Barros, M.J.; Maravilhas-Lopes, S. A Business Intelligence Maturity Evaluation Model for Management Information Systems Departments. In *Handbook of Research on Emerging Technologies for Effective Project Management*; IGI Global: Hershey, PA, USA, 2019. [CrossRef]
49. Demchenko, Y.; Manieri, A.; Belloum, A. EDISON Data Science Framework (EDSF) | Edison Project', Book. Available online: https://edison-project.eu/edison/edison-data-science-framework-edsf/ (accessed on 28 December 2024).
50. United Nations Conference on Trade and Development (UNCTAD). UNCTAD Handbook of Statistics 2020—Total Merchandise Trade. Available online: https://unctadstat.unctad.org/insights/theme/11 (accessed on 29 March 2023).
51. Flath, C.M.; Stein, N. Towards a data science toolbox for industrial analytics applications. *Comput. Ind.* **2018**, *94*, 16–25. [CrossRef]
52. Lasi, H.; Fettke, P.; Kemper, H.G.; Feld, T.; Hoffmann, M. Industry 4.0. *Bus. Inf. Syst. Eng.* **2014**, *6*, 239–242. [CrossRef]
53. Behun, M.; Gavurova, B.; Tkacova, A.; Kotaskova, A. The Impact of The Manufacturing Industry on the Economic Cycle of European Union Countries. *J. Compet.* **2018**, *10*, 23–39. [CrossRef]
54. European Commission. *Industry in Europe—Facts & Figures on Competitiveness & Innovation*; European Commission: Brussels, Belgium, 2017. [CrossRef]
55. Deloitte. Industry 4.0 Challenges and Solutions for the Digital Transformation and Use of Exponential Technologies. Available online: https://www.pac.gr/bcm/uploads/industry-4-0-deloitte-study.pdf (accessed on 29 March 2021).
56. I Scoop Europe. Industry 4.0: Fourth Industrial Revolution Guide to Industrie 4.0. Available online: https://www.i-scoop.eu/industry-4-0/ (accessed on 31 March 2021).
57. Hernandez-de-Menendez, M.; Morales-Menendez, R.; Escobar, C.A.; McGovern, M. Competencies for Industry 4.0. *Int. J. Interact. Des. Manuf.* **2020**, *14*, 1511–1524. [CrossRef]
58. Ghobakhloo, M.; Iranmanesh, M.; Foroughi, B.; Tseng, M.L.; Nikbin, D.; Khanfar, A.A.A. Industry 4.0 digital transformation and opportunities for supply chain resilience: A comprehensive review and a strategic roadmap. *Prod. Plan. Control.* **2023**. [CrossRef]
59. Vallerio, M.; Perez-Galvan, C.; Navarro-Brull, F.J. Industrial Data Science for Batch Manufacturing. *Comput. Aided Chem. Eng.* **2024**, *53*, 2965–2970. [CrossRef]

60. Reinsel, D.; Gantz, J.; Rydning, J. *The Digitization of the World from Edge to Core*; International Data Corporation: Framingham, MA, USA, 2018.
61. Alcácer, V.; Cruz-Machado, V. *Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems*; Elsevier: Amsterdam, The Netherlands, 2019. [CrossRef]
62. Baltassis, E.; Coulin, A.-D.; Gourévitch, A.; Khendek, Y.; Quarta, L. *A Rough Road to Data Maturity*; Boston Consulting Group: Hong Kong, China, 2019.
63. Oussous, A.; Benjelloun, F.Z.; Lahcen, A.A.; Belfkih, S. Big Data technologies: A survey. *J. King Saud Univ. Comput. Inf. Sci.* **2018**, *30*, 431–448. [CrossRef]
64. Zhang, L. Big Data Analytics for Fault Detection and its Application in Maintenance. Ph.D. Thesis, Luleå University of Technology, Luleå, Sweden, 2014.
65. Sharma, A.; Pandey, H. Big Data and Analytics in Industry 4.0. In *A Roadmap to Industry 4.0: Smart Production, Sharp Business and Sustainable Development*; Nayyar, A., Kumar, A., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 57–72. [CrossRef]
66. Castelo-Branco, I.; Cruz-Jesus, F.; Oliveira, T. Assessing Industry 4.0 readiness in manufacturing: Evidence for the European Union. *Comput. Ind.* **2019**, *107*, 22–32. [CrossRef]
67. Upadhyaya, M. Impact of big data on analytics—Capgemini. In *Nasscom Big Data and Analytics Summit*; Capgemini: Hong Kong, China, 2014.
68. Holder, S. *The Analytics Continuum*; SAS: Stockholm, Sweden, 2018.
69. Ghiraldini, M. Multi-Touch Attribution, Where Do I Start? Available online: https://blog.math.group/multi-touch-attribution-por-onde-devo-comecar (accessed on 5 June 2020).
70. Król, K.; Zdonek, D. Analytics Maturity Models: An Overview. *Information* **2020**, *11*, 142. [CrossRef]
71. Grossman, R.L. A framework for evaluating the analytic maturity of an organization. *Int. J. Inf. Manag.* **2018**, *38*, 45–51. [CrossRef]
72. Thomas, D.M. A Survey on the Evolution of Data Analytics and the Future in Data Science. *Int. J. Innov. Sci. Res. Technol.* **2020**, *5*, 1699–1703. Available online: https://ijisrt.com/assets/upload/files/IJISRT20AUG827.pdf (accessed on 31 March 2021).
73. Boukouvala, F.; Li, J.; Xiao, X.; Floudas, C.A. Data-driven modeling and global optimization of industrial-scale petrochemical planning operations. *Proc. Am. Control Conf.* **2016**, *2016*, 3340–3345. [CrossRef]
74. Ge, Z. Review on data-driven modeling and monitoring for plant-wide industrial processes. *Chemom. Intell. Lab. Syst.* **2017**, *171*, 16–25. [CrossRef]
75. Risse, M. *Using Data Analytics to Improve Operations and Maintenance*; Power Mag—Seeq Corp.: Seattle, DC, USA, 2018.
76. Karim, R.; Westerberg, J.; Galar, D.; Kumar, U. Maintenance Analytics—The New Know in Maintenance. *IFAC-PapersOnLine* **2016**, *49*, 214–219. [CrossRef]
77. Kumar, V.; Garg, M.L. Predictive Analytics: A Review of Trends and Techniques. *Artic. Int. J. Comput. Appl.* **2018**, *182*, 975–8887. [CrossRef]
78. Williamson, B. Digital education governance: Data visualization, predictive analytics, and 'real-time' policy instruments. *J. Educ. Policy* **2015**, *31*, 123–141. [CrossRef]
79. Jiang, Y.; Yin, S.; Kaynak, O. Data-driven monitoring and safety control of industrial cyber-physical systems: Basics and beyond. *IEEE Access* **2018**, *6*, 47374–47384. [CrossRef]
80. Wang, J.; Zhang, W.; Shi, Y.; Duan, S.; Liu, J. Industrial Big Data Analytics: Challenges, Methodologies, and Applications. *IEEE Trans. Autom. Sci. Eng.* **2018**. [CrossRef]
81. Fisher, O.J.; Watson, N.J.; Escrig, J.E.; Witt, R.; Porcu, L.; Bacon, D.; Rigley, M.; Gomes, R.L. Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems. *Comput. Chem. Eng.* **2020**, *140*, 106881. [CrossRef]
82. Zaman, I.; Pazouki, K.; Norman, R.; Younessi, S.; Coleman, S. Challenges and opportunities of big data analytics for upcoming regulations and future transformation of the shipping industry. *Procedia Eng.* **2017**, *194*, 537–544. [CrossRef]
83. Sheriff, S. 8 Major Challenges Faced by Data Scientists—Acuvate. Available online: https://acuvate.com/blog/challenges-faced-by-data-scientists/ (accessed on 5 December 2021).
84. Bradford, L. 8 Real Challenges Data Scientists Face | Forbes. *Forbes*, 6 September 2018. Available online: https://www.forbes.com/sites/laurencebradford/2018/09/06/8-real-challenges-data-scientists-face/ (accessed on 15 December 2020).
85. Waters, R. How machine learning creates new professions—And problems | Financial Times. *Financial Times*, 29 November 2017.
86. Wani, M.A.; Jabin, S. Big data: Issues, challenges, and techniques in business intelligence. *Adv. Intell. Syst. Comput.* **2018**, *654*, 613–628. [CrossRef]
87. Dzemyda, G.; Bernatavičienė, J.; Kacprzyk, J. Data Science: New Issues, Challenges and Applications. In *Studies in Computational Intelligence*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; Volume 869. [CrossRef]

88. Kaggle. 2017 Kaggle ML & DS Survey—A Big Picture View of the State of Data Science and Machine Learning. Available online: https://www.kaggle.com/kaggle/kaggle-survey-2017 (accessed on 4 December 2020).
89. Hayes, B. Top 10 Challenges to Practicing Data Science at Work | Business over Broadway—Analytics, Data Science, Machine Learning. Available online: http://businessoverbroadway.com/2018/03/18/top-10-challenges-to-practicing-data-science-at-work/ (accessed on 4 December 2020).
90. Redman, T.C. Data quality must reads for researchers. *J. Decis. Syst.* **2017**, *26*, 203–206. [CrossRef]
91. 6 Pillars of Data Quality and How to Improve Your Data—IBM Blog. Available online: https://www.ibm.com/blog/6-pillars-of-data-quality-and-how-to-improve-your-data/ (accessed on 3 July 2024).
92. Jaya, M.A.J.M.I.; Sidi, F.; Ishak, I.; Affendey, L.S. A review of data quality research in achieving high data quality within organization. *J. Theor. Appl. Inf. Technol.* **2017**, *95*, 2647–2657.
93. Cai, L.; Zhu, Y. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Sci. J.* **2015**, *14*, 2. [CrossRef]
94. Inkinen, T.; Helminen, R.; Saarikoski, J. Port Digitalization with Open Data: Challenges, Opportunities, and Integrations. *J. Open Innov. Technol. Mark. Complex.* **2019**, *5*, 30. [CrossRef]
95. Lu, Y. Industry 4.0: A survey on technologies, applications and open research issues. *J. Ind. Inf. Integr.* **2017**, *6*, 1–10. [CrossRef]
96. McDonnell, C. Measuring Data Quality | Deloitte Ireland | Tax, Deloitte Private. Available online: https://www.deloitte.com/ie/en/services/risk-advisory/research/measuring-data-quality.html (accessed on 7 March 2021).
97. Syberg, M.; West, N.; Schwenken, J.; Adams, R.; Deuse, J. A requirement-driven approach for competency-based collaboration in industrial data science projects. *Int. J. Prod. Manag. Eng.* **2024**, *12*, 79–90. [CrossRef]
98. Umachandran, K. *Industry 4.0—A Brief*. 2020. Available online: https://www.researchgate.net/publication/342734177_Industry_40_-_A_brief (accessed on 21 March 2024).
99. Pires, I.M.; Garcia, N.M.; Pombo, N.; Flórez-Revuelta, F. From data acquisition to data fusion: A comprehensive review and a roadmap for the identification of activities of daily living using mobile devices. *Sensors* **2016**, *16*, 184. [CrossRef] [PubMed]
100. Horton, N.J.; Baumer, B.S.; Wickham, H. Setting the stage for data science: Integration of data management skills in introductory and second courses in statistics. *arXiv* **2015**, arXiv:1502.00318 2015.
101. Lohr, S. 'For Big-Data Scientists, "Janitor Work" Is Key Hurdle to Insights. *The New York Times*, 18 August 2014. Available online: https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html (accessed on 2 August 2024).
102. Wickham, H. Tidy data. *J. Stat. Softw.* **2014**, *59*, 1–23. [CrossRef]
103. Accenture. *The Human Impact of Data Literacy the Human Impact of Data Literacy Report*; Accenture: Dublin, Ireland, 2020.
104. Sarker, I.H. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *Sn Comput. Sci.* **2021**, *2*, 377. [CrossRef] [PubMed]
105. Grus, J. *Data Science from Scratch*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2015; Volume 1542.
106. Bhattarai, B.P.; Paudyal, S.; Luo, Y.; Mohanpurkar, M.; Cheung, K.; Tonkoski, R.; Hovsapian, R.; Myers, K.S.; Zhang, R.; Zhao, P.; et al. Big data analytics in smart grids: State-of-the-art, challenges, opportunities, and future directions. *IET Smart Grid* **2019**, *2*, 141–154. [CrossRef]
107. Mordor Intelligence. *Data Governance Market—Growth, Trends, and Forecast (2020—2025)*; Mordor Intelligence: Nanakramguda, India, 2020.
108. Gandomi, A.; Haider, M. Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manag.* **2015**, *35*, 137–144. [CrossRef]
109. Bjørn-andersen, N. Ships and Port Idle Time: Who are the Culprits? August 2018. Available online: https://www.academia.edu/119753979/Ships_and_Port_Idle_Time_Who_are_the_Culprits (accessed on 2 November 2024).
110. Sarriera, J.M.; Araya, G.; Serebrisky, T.; Briceno-Garmendia, C.; Schwartz, J. *Benchmarking Container Port Technical Efficiency in Latin America and the Caribbean: A Stochastic Frontier Analysis*; World Bank: Washington, DC, USA, 2013.
111. Farooq, S.U.; Quadri, S.M.K.; Ahmad, N. Metrics, models and measurements in software reliability. In Proceedings of the IEEE 10th Jubilee International Symposium on Applied Machine Intelligence and Informatics, SAMI 2012, Herl'any, Slovakia, 26–28 January 2012; pp. 441–449. [CrossRef]
112. Chen, C.L.P.; Zhang, C.Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Inf. Sci.* **2014**, *275*, 314–347. [CrossRef]
113. Rathod, A. Pattern Recognition in AI: How It Works and Applications. AI Perceiver. Available online: https://aiperceiver.com/what-is-pattern-recognition-in-artificial-intelligence/ (accessed on 2 November 2024).
114. Keilis-Borok and, A. Soloviev. Pattern Recognition Methods and Algorithms. In *Ninth Workshop on Non-linear Dynamics and Earthquake Predictions*; The Abdus Salam International Centre for Theoretical Physics: Trieste, Italy, 2007; pp. 1–21. Available online: https://indico.ictp.it/event/a06219/session/53/contribution/26/material/0/0.pdf (accessed on 25 September 2024).

115. Casasent, D. Pattern Recognition: A Review. *IEEE Spectr.* **1981**, *18*, 28–33. [CrossRef]
116. Awasthi, A.; Krpalkova, L.; Walsh, J. Deep Learning-Based Boolean, Time Series, Error Detection, and Predictive Analysis in Container Crane Operations. *Algorithms* **2024**, *17*, 333. [CrossRef]