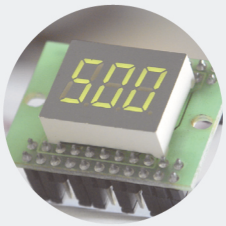
INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

MESTRADO EM ENGENHARIA ELECTROTÉCNICA E DE COMPUTADORES

O Impacto de Big Data Analytics na Gestão das Cadeias de Abastecimento

JOÃO ANDRÉ MARTINS E SILVA

julho de 2023

POLITÉCNICO
DO PORTO

Politécnico do Porto
Instituto Superior de Engenharia do Porto

The Impact of Big Data Analytics on Supply Chain Management

João André Silva

Master in Electrical and Computer Engineering
Specialization Area of Systems And Industrial Planning

Departamento de Engenharia Eletrotécnica
Instituto Superior de Engenharia do Porto

July, 2023

This dissertation partially satisfies the requirements of the Thesis/Dissertation course of the program Master in Electrical and Computer Engineering, Specialization Area of Systems And Industrial Planning.

Candidate: João André Silva, No. 1170747, `1170747@isep.ipp.pt`

Scientific Guidance: António Amaral, `sal@isep.ipp.pt`

Scientific Co-Guidance: Tiago Pinho, `tiago.pinho@esce.ips.pt`

Departamento de Engenharia Eletrotécnica
Instituto Superior de Engenharia do Porto
Rua Dr. António Bernardino de Almeida, 431, 4200–072 Porto

July, 2023

Abstract

This research attempts to understand better how Big Data Analytics affects modern businesses and how it could impact daily operations in the global supply chain and consumer engagement activities. In order to emphasise the most influential scholars and the range of research topics, a literature review was carried out to reveal the state of the art in this field. The results show that many businesses are already utilising Big Data Analytics to gain a competitive edge by spotting patterns and trends to help them make more informed and robust decisions. Several Machine Learning and Artificial Intelligence approaches are also being implemented to automate the collection of data and the ability to process it in real-time. Notwithstanding, there is still a significant research gap about data quality challenges and how to resolve them, as there are still not enough real-world applications to adequately illustrate the benefits of these technologies in ordinary supply chain operations. This research outlines the need for a more in-depth analysis of how Big Data Analytics can be used to restructure a company's supply chain architecture and provides examples of how Big Data and related technologies can aid businesses in boosting productivity and improving cost-effectiveness. Towards whether these assertions could be implemented in real case scenarios or if the practical reality differed significantly from the theoretical one, all these propositions were explored by interviewing several supply chain professionals using a preset of questions. The key findings from these analyses were then provided at the conclusion of this document.

Keywords: Big Data Analytics; Supply chain; Data quality; Artificial Intelligence; Machine Learning

Resumo

Esta pesquisa tenta entender como Big Data Analytics afeta as empresas modernas e como isso pode impactar as operações diárias na cadeia de abastecimento global e as atividades de relacionamento direto com o consumidor. A fim de enfatizar os investigadores mais influentes e a variedade de tópicos de pesquisa, foi realizada uma revisão da literatura para revelar o estado da arte neste campo. Os resultados mostram que muitas empresas já utilizam Big Data Analytics para ganhar uma vantagem competitiva, identificando padrões e tendências que os poderão ajudar a tomar decisões mais informadas e assertivas. Várias abordagens de *Machine Learning* e Inteligência Artificial estão também a ser implementadas para automatizar a obtenção de dados e a capacidade de processá-los em tempo real. No entanto, existe ainda uma lacuna significativa de pesquisa sobre os desafios de qualidade dos dados e de como resolvê-los, uma vez que ainda não existem aplicações reais suficientes para ilustrar adequadamente os benefícios destas tecnologias nas operações normais da cadeia de abastecimento. Esta pesquisa descreve a necessidade de uma análise mais aprofundada de como Big Data Analytics pode ser usado para reestruturar a arquitetura da cadeia de abastecimento de uma empresa e fornece exemplos de como Big Data e todas as tecnologias relacionadas podem ajudar as empresas a aumentar a produtividade e melhorar a sua rentabilidade. Para saber se estas afirmações poderiam ser implementadas em cenários de caso real ou se a realidade prática diferia significativamente da teórica, todas as proposições foram exploradas através de entrevistas com vários profissionais da área Logística, usando um conjunto predefinido de perguntas. As principais conclusões destas análises foram então apresentadas na conclusão deste documento.

Palavras-Chave: Big Data Analytics; Cadeia de abastecimento; Qualidade da informação; Inteligência Artificial; Machine Learning

Contents

List of Figures

List of Tables

List of Acronyms

AI	Artificial Intelligence
B2B	Business to Business
B2C	Business to Consumer
BD	Big Data
BDA	Big Data Analytics
C.E.O.	Chief Executive Officer
C.O.O.	Chief Operating Officer
CRM	Customer Relationship Management
ERP	Enterprise Resource Planning
ETL	Extract, Transform and Load
GPS	Global Positioning System
IoT	Internet of Things
IT	Information Technology
JIT	Just in Time
KPI	Key Performance Indicator
ML	Machine Learning
PA	Predictive Analytics
RFID	Radio frequency identification
SCM	Supply Chain Management
SLR	Systematic Literature Review
WMS	Warehouse Management System

Chapter 1

Introduction

1.1 Supply Chain Context

The evolution of Supply Chain Management (SCM) has been marked by an increasing degree of integration of several distinct operations. From the implementation of new handling procedures to reliance on ocean-going vessels, from containerisation to computerisation, global supply chains are more complex and efficient than ever before [1][2].

Due to a lack of broader transportation choices and the high expense of transferring materials between regions before the 20th century, all supply chain elements were kept essentially local [3]. As countries' resources expanded through industrial revolutions, moving cargo over longer distances became faster, simpler, and less expensive (although supply chains still tended to be limited to countries) [4].

The development of mass manufacturing along assembly lines in the late 1920s created the groundwork for SCM. The concept of creating similar items on a big scale with enhanced efficiency permanently transformed commerce and supply networks [5]. The Supply Chain Management evolves, in conjunction with the use of pallets in warehouses, enabling items to be stacked vertically, conserving space and increasing the efficiency of goods handling [6]. The introduction of shipping containers was arguably one of the most significant changes in global supply networks. Containerisation, or container transportation, increased not only the amount of accessible space for products but also the speed of freight movements while cutting the overall transportation cost [7]. The increased speed was due to more efficient storage procedures as well as the transport terminal efficiency. The refinement of this logistic procedure, which included loading and unloading commodities, marked a new era of worldwide trade [8].

Computerisation also gained popularity in the second part of the 20th century [9]. Logistic records and data were initially collected, transmitted, and reported on paper. Data computerisation began to streamline logistics by developing optimisation models and algorithms that could predict operational issues in a supply chain, opening up opportunities in a variety of areas such as more accurate forecasting, better warehouse storage, truck routing, and better inventory management [10]. The introduction of solutions such as Enterprise Resource Planning (ERP) systems enabled businesses to use software to handle all of their activities, including automating corporate tasks, centralising information, managing finances, and measuring performance, therefore changing the supply chain to a global model [11][12].

The emergence of such information systems brought context to the use of some analytics, adding a monitoring dimension to SCM [13]. As product life cycles have shortened and efficiency has grown, SCM had to rely on technology to satisfy stakeholders' expectations. Namely, requesting real-time monitoring has risen in importance as public pressure to stay sustainable and socially responsible has increased [14][15]. However, combining business models and supply chain innovation while leveraging fast improvements in Information Technology (IT), particularly Big Data (BD) and small data capabilities, is a unique challenge for supply network designers [16][17].

Big Data Analytics (BDA) describes the process of identifying trends, patterns, and correlations in vast quantities of raw data to help make data-informed decisions [18]. These processes employ well-known statistical analysis techniques, such as clustering and regression, and apply them to large business datasets [19]. Data analytics is transforming how companies interact and redesign their supply chain processes, as valuable insights can be obtained through data analysis, enabling the industry players to streamline factory functions, optimise routing, reduce costs, and ensure the overall transparency of the supply chain [18]. In addition, the capacity to track items, people, and consumption patterns at a lower level could provide insights beyond product replenishment, particularly into consumer behaviour, product quality, and physical location [20].

Historically, supply networks were not designed to be transparent, especially because complete exposure could jeopardise the information's confidentiality held by competitive providers [21][22]. They may learn that they do not have to be as competitive if they know that specific criteria (e.g., price) are given a low weighting. However, with the attention placed on firms' sourcing processes, supply chain transparency has become a must in many industries [21][23]. Consumers are increasingly interested in where, how, and by whom the items they buy are manufactured.

Meanwhile, governments, non-governmental organisations, and other stakeholders are putting pressure on businesses to be more transparent by rating and advertising their transparency levels, and providing more information about their supply chains behaviour [24][25].

In reality, building a transparent supply chain entails more than deciding what information to provide to customers, as firms must first get insight into their supply networks [26][27]. However, the work and resources required to monitor first-tier suppliers, much alone upper-tier suppliers, can be costly and time consuming [28]. Deeper insight into the whole supply network could be attained by allowing firms to swiftly and efficiently harness massive amounts of data from different sources while using BDA [29]. The broader assessment of Big Data can increase visibility throughout the supply chain, giving organisations a further advantage by managing risk accordingly and supporting their decision making process [30]. Therefore, supply chain analytics will enable the translation of a multitude of information into choices. Because managers can simulate disturbances and make vital, timely choices, real-time data translates into organisational or supply chain resilience and flexibility [31].

1.2 Objectives

Over the past few years, researchers have been trying to fully understand these technologies' implications on multiple business metrics and overall supply chain performance [32][33]. With the introduction of these newer methodologies in data-rich environments, some traditional procedures may have turnout obsolete, considering the large volume of data flowing between each supply chain participant [34][35]. Even though implementing innovative techniques to manage BDA appears to be the ideal way to propel enterprises to the next level, many business partners are still hesitant to change their company's culture due to a lack of knowledge about BD [29][36]. The primary objective of this research is to explore the implications of BDA in Supply Chain Management processes through an extensive literature review that discusses the applications of BD and its associated challenges, providing the logistics and SCM community with fundamental research on how organisations can exploit BDA capabilities to provide logistics and supply chain insights. Secondly, the literature findings will then be explored through several interviews with supply chain and logistics professionals to gain a better understanding of which are the upcoming changes in the global supply chain, as well as the main differences between the perspectives available in the literature and within the Supply Chain Management practice community.

1.3 Research Design

The research design follows the research onion framework proposed by Saunders et al. [37]. The Research Onion consists of six main layers, each describing a stage of the research process. These layers focus on (i) Research philosophy; (ii) Approach to theory development; (iii) Methodological choice; (iv) Strategy; (v) Time horizon; and (vi) Techniques and procedures. The philosophical approach followed was critical realism which recognises the importance of studies incorporating multiple evaluation perspectives like the individual, the group, or the organisation. This philosophy sees experiences and observations as a result of objective structures. The research approach followed was inductive, with the necessity of deeply understanding the context in which the investigation takes place, its particularities, and assumptions. A new theory could be developed according to the data collected and assessed. The methodological choice followed was qualitative, focused on examining descriptive data that cannot be numerically measured, such as personal experiences or other descriptions of empirical nature. Therefore, the research purpose is exploratory, in which structured interviews are conducted with experts to develop more profound knowledge within our specific research scope. The research strategy is based on selecting the methods used to obtain primary data. A considerable number of interviews were performed with experts with different experiences and roles within the SCM area to capture the idiosyncrasies. The time horizon selected for this research's purposes is cross-sectional because the objective of this research was to capture the participants' perspectives during the master thesis development. The primary data was collected during the semi-structured interviews performed with the experts that were transcribed in the end.

1.4 Dissertation Structure

This dissertation is divided into three further sections in addition to this introduction. The methodology utilised during the literature review is discussed in Chapter 2, along with the evolution of these topics over the past ten years, the kinds of documents found in the data set utilised, and the most well-known authors and sources where these kinds of articles are published. The literature review is also presented in this section, confirming which publications were thoroughly examined and what state-of-the-art findings were made. Multiple interviews were undertaken to compare these findings with real-world examples and see whether the theoretical foundations of these applications, which various scholars proposed, would be effective in some situations. The results of these interviews, a brief overview of each interviewee and the methodology employed for this qualitative research task, were reported in Chapter 3. As a conclusion to this work and a suggestion for further research, Chapter 4 analyses the implications and limits of this dissertation.

Chapter 2

Literature Review

A literature review serves as the foundation for high-quality Supply Chain Management (SCM) research, maximising the study paper's relevance, originality, and impact [38]. Understanding the existing literature on a given topic is essential for all phases of a research project. This research project adopted a Systematic Literature Review (SLR) to analyse state of the art available on this subject matter. A systematic review based on a clearly stated issue, discovers relevant research, assesses its quality, and summarises the evidence using specified methods. Thus, the presented research was divided into five sections:

1. **Framing questions for a review:** The issues that will be addressed by the review, such as, "What are the primary benefits of utilising Big Data Analytics in Supply Chain processes?", "How can Big Data and new security technologies improve inter-company collaboration?", and "How can Big Data improve visibility and traceability within modern supply chains?"

2. **Identifying relevant publications:** To collect as many relevant citations as possible, a wide range of Supply Chain Management (SCM) and logistics databases were searched to locate firsthand research on the impact of Big Data Analytics (BDA) in business processes.

3. **Assessing the quality of studies:** The assessment of research quality is critical for constructing a solid and well-founded review. Selected studies should be subjected to a more refined quality assessment process using metrics such

as the number of citations, the overall characterisation of the author (main research areas, average number of citations, number of published papers), and the journal in which the paper was published (and its rating).

4. **Summarising the evidence:** Review of the collected information, extracting the key concepts and takeaways from each publication.

5. **Interpreting the findings:** Evaluate the findings in light of the many forms of study, such as case studies, future research suggestions, and refutations of previous studies.

2.1 Bibliometric Analysis

A systematic bibliometric analysis conducted using two well-known scientific databases, Web of Science and Scopus. These searches were performed using "big data analytics" and "supply chain analytics" as entry points. Given the recent developments on this topic, only English language-based articles published in the last decade were considered. Due to using these parameters, the search returned almost 1300 documents.

To further narrow down the number of publications of relevance on this matter, only papers with fifty or more citations were chosen to separate the most established writers from those who are only now beginning to share their findings. Figure 2.1 illustrates the progression of state-of-the-art articles available in scientific databases throughout the study area.

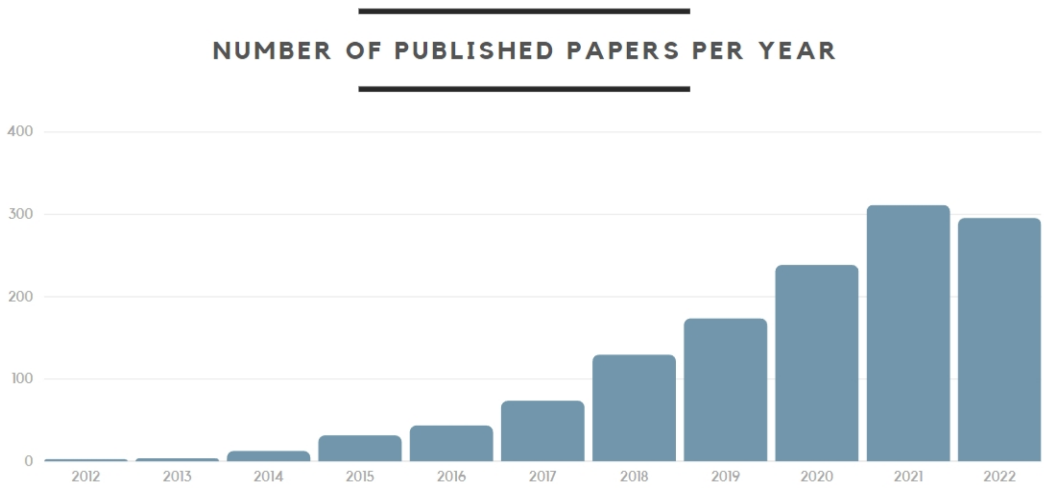


Figure 2.1: Evolution of scientific papers published across the last decade.

Given the large number of documents identified, it is also essential to determine what kind of papers were published on this subject.

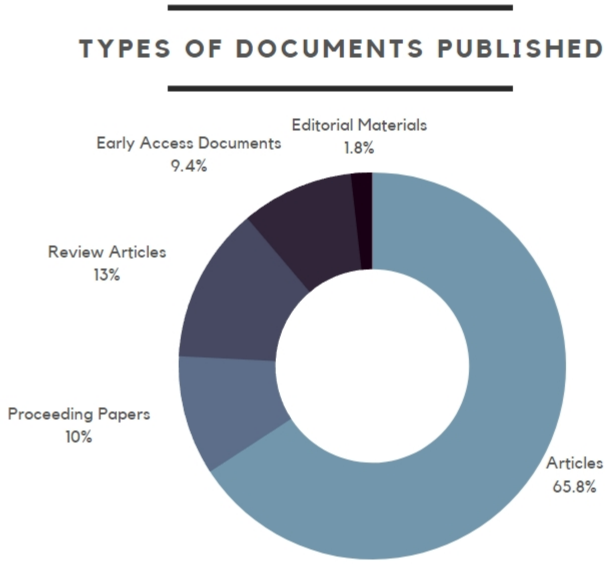


Figure 2.2: Types of documents considered for analysis.

According to the ring chart, articles constitute more than 60% of the documents saved in the database. This outcome was initially anticipated since the general subject is not entirely new but a combination of multiple research fields, such as Data Analytics, Artificial Intelligence, Machine Learning, and Supply Chain Management. Review papers make up a small percentage of the total number of publications since they give an in-depth review of the information available from earlier studies and can propose possible areas of inquiry to look into next while still concluding the current data. Proceedings papers in journals represent just a tiny portion of all documents collected. These publications have a format similar to conventional journal articles. However, they go through a less demanding review process and are published faster with a lesser scientific impact (citation level).

2.1.1 Author Analysis

Since this subject matter was considered a hot topic at recent Supply Chain and logistics conferences, the number of researchers publishing scientific papers about it has been increasingly high over the past few years. A network map with the most influential authors of this subject was created by analysing the selected paper data set with the software VOSViewer.

This analysis only considered writers whose papers had 100 or more citations to filter the most established authors from the ones who are just now starting to share their research.

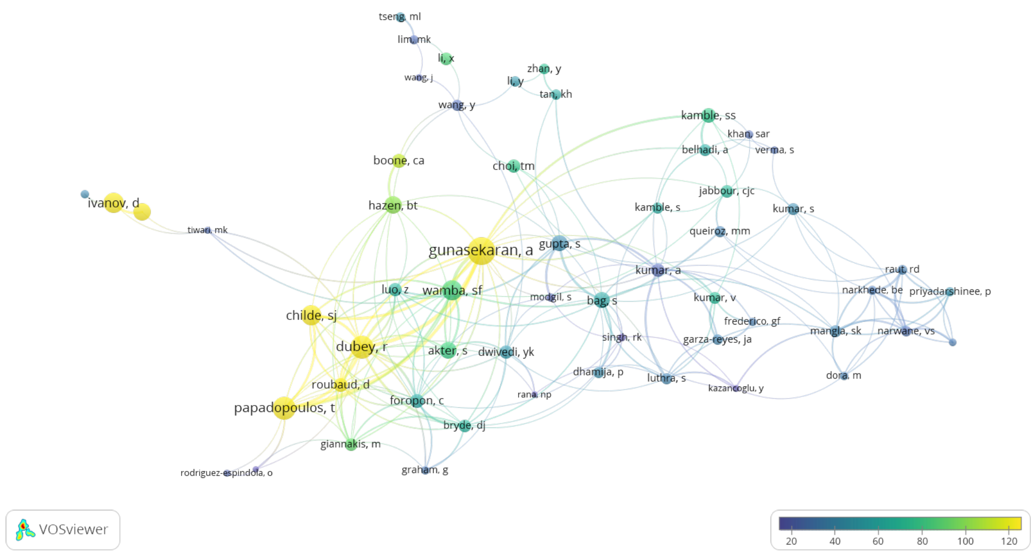


Figure 2.3: Most relevant authors according to their number of citations.

As presented in Figure 2.3, authors like Angappa Gunasekaran, Thanos Papadopoulos, Rameshwar Dubey, and Dmitry Ivanov are well established researchers on this study theme, as their papers collect over 100 citations on average, even though they are not necessarily the scientists with the most articles published. Most of these authors work directly with each other in multiple papers since many reside in the same region. The number of links between each cluster in the previous figure represents this work chemistry. Papers published in the USA are undoubtedly the most cited throughout the years, with the People's Republic of China, Germany, and England coming next to them.

Since the data set contains multiple documents from the last decade, it is also appropriate to look into each author's average publication date to determine whether the high number of citations is due to them actively publishing high-quality studies or are all related to one unique article.

As shown in Figure 2.1, it is unlikely to have a large cluster in the first years of the data set because information on this issue was still scarce. Most of the articles developed simply discussed future and prospective applications of Big Data in Supply Chain activities.

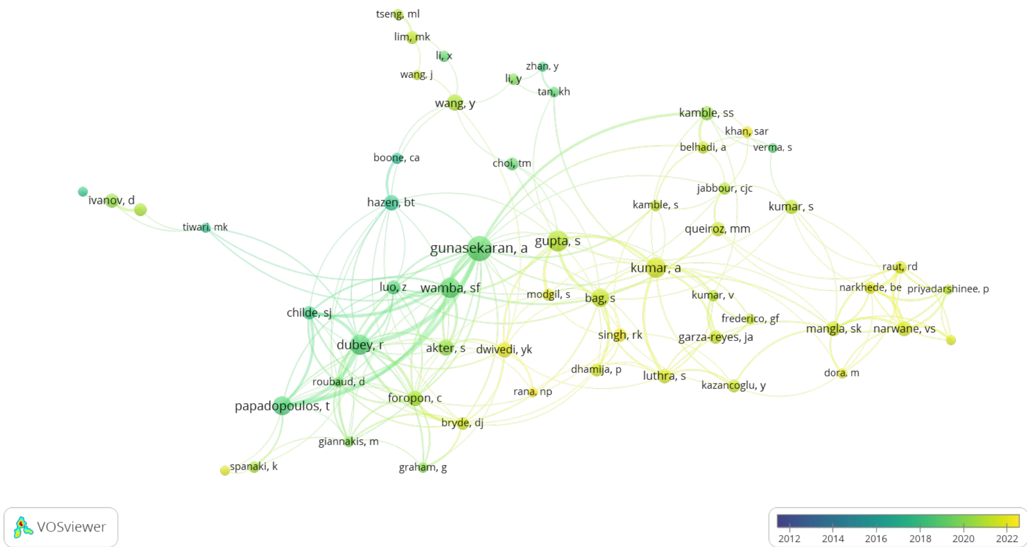


Figure 2.4: Average publication year of each author.

The previous figure clearly illustrates two major historical turning points during the last decade: The first cluster of highly cited documents dates back to 2016, when authors began to create studies by swiftly analysing Big Data Analytics applications at the time, providing the scientific community with new ideas and research proposals that could take place soon, on more developed companies; the second cluster is centred on 2022, and writers rely their study on previous publications (from the first cluster) to postulate about the possible application of the theoretical concepts mentioned in those studies, in modern Supply Chains.

It is also essential to fully understand the difference between concepts covered in papers published in 2016 and those published in 2022. Since Supply Chain concerns evolve rapidly, the paradigm of technology and digital process in Supply Chain Management activities have shifted dramatically, requiring new researchers to develop papers from a more technological standpoint, discussing automation and process optimisations. In contrast, papers published a few years ago would have focused more on creating unified process frameworks for risk reduction and quality control.

A brief review of each paper's primary subject matter was done, contrasting critical points made in older articles with newer ones:

Table 2.1: Comparison between key concepts on papers from 2016 and 2022.

Papers from 2016	**Papers from 2022**
Big Data Analytics	Machine Learning
Model	Performance
Quality	Optimisation
Risk	Challenges

2.1.2 Journal Analysis

Besides exploring the relevance of each author and the number of papers published, it is also necessary to analyse where the articles were issued since articles published in high-quality journals are usually peer-reviewed and oriented towards the scientific community. Most of these journals are ranked according to their quartile, which considers the journal's impact factor, average number of citations, and index value.

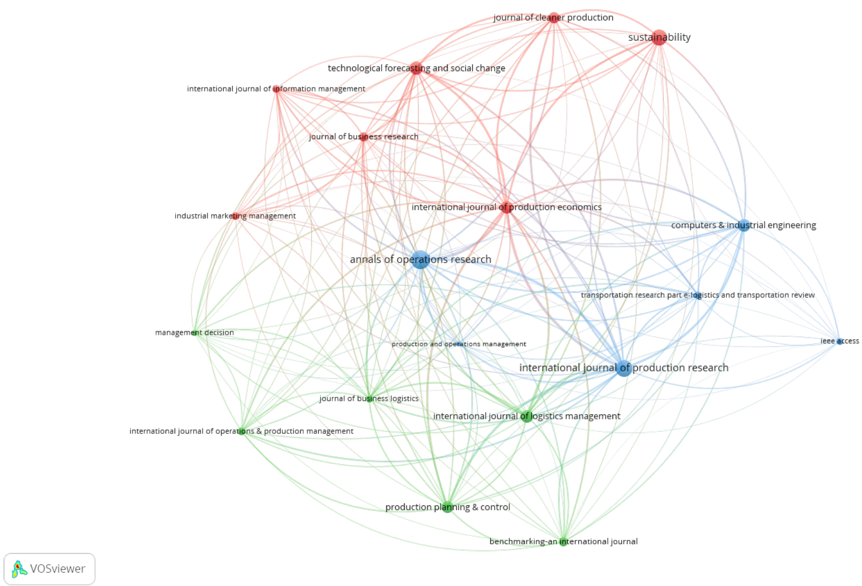


Figure 2.5: Most relevant sources according to their number of citations.

The previous network demonstrates the wide range of scholarly outlets that have distributed and published research articles on this topic. A high volume of papers is published annually in journals like Annals of Operations Research, International Journal of Production Research, and Sustainability. In contrast, journals like the International Journal of Production Economics, Journal of Business Logistics, and Journal of Cleaner Production typically publish fewer but highly cited documents yearly (since they focus more on those industries).

To further understand the notability of each journal, a brief analysis of the ranking of each periodical was made, taking into account the top 5 journals with the most citations. This analysis will help distinguish the most relevant papers from the not-so-relevant ones, favouring a better Literature Review.

Table 2.2: Ranking analysis of the leading scientific journals used.

Journal	**Study Areas**	**Quartile**	**H-Index**
International Journal of Production Economics	Management, Accounting Decision Sciences Economics, Econometrics	Q1	197
International Journal of Production Research	Management, Accounting Decision Sciences Manufacturing Engineering	Q1	153
Annals of Operations Research	Decision Sciences Management Sciences	Q1	111
Journal of Cleaner Production	Management, Accounting Manufacturing Engineering Environmental Science Sustainability	Q1	232
Computers & Industrial Engineering	Computer Science Engineering	Q1	136

Most sources utilised to consult papers review frequently publish studies about logistical domains, such as Management and Accounting or Manufacturing Engineering. However, several valuable publications in journals specialise in specific concepts, such as Sustainability and Green Supply Chains, Computer Science, Industry Economics, and Forecasting.

This analysis allowed the selection of the most pertinent publications on this subject, leading to a more thorough investigation and comprehension of their key goals. Fifty documents published on previously mentioned sources were chosen and reviewed to achieve this.

2.2 Papers of Interest

Information on all of the studies used for this research is presented in table 2.3, including the title, authors, publication year, number of citations (as of the moment of writing, according to Web of Science), and the paper's keywords. The subject matter of each publication can also give the scientific community some insight into what the research is covering, whether it be theoretical assumptions or a real-world case study on that particular subject. The majority of the publications assessed were theoretical studies that provided some insight into current research on each issue while providing space for future discussion between existing literature and future research recommendations. This paradigm has shifted in recent years, with several papers now actually implementing previous theories to assess their usefulness in real-world applications.

Table 2.3: List of documents reviewed.

Title	Authors	Year	Citations	Keywords
Beyond the hype: Big data concepts, methods, and analytics	Gandomi, A.; Haider, M.	2015	1815	Big data analytics; Big data definition; Unstructured data analytics; Predictive analytics
Intelligent Manufacturing in the Context of Industry 4.0: A Review	Zhong, R.; Xu, X.; Klotz, E.; Newman, S.	2017	1116	Intelligent manufacturing; Industry 4.0; Internet of Things; Manufacturing systems; Cloud manufacturing; Cyber-physical system
Industry 4.0 technologies: Implementation patterns in manufacturing companies	Frank, A.; Dalenogare, L.; Ayala, N.	2019	852	Industry 4.0; Smart Manufacturing; Digital transformation; Manufacturing companies
Sustainable supply chain management: evolution and future directions	Carter, C.; Easton, P.	2011	836	Supply chain management; Economic sustainability; Social responsibility; Environmental management; Economic performance

Title	Authors	Year	Citations	Keywords
Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management	Waller, M.; Fawcett, S.	2013	657	Data science; Predictive analytics; Big data; Logistics; Supply chain management; Design; Collaboration; Integration; Education
Big data analytics in logistics and supply chain management: Certain investigations for research and applications	Wang, G.; Gunasekaran, A.; Ngai, E.; Papadopoulos, T.	2016	642	Big data; Supply chain analytics; Maturity model; Holistic business analytics; Methodologies and techniques
Literature review of Industry 4.0 and related technologies	Oztemel, E.; Gursev S.	2020	628	Industry 40; Smart factory; Internet of things; Cyber-physical systems; Cloud systems; Big data
Toward Scalable Systems for Big Data Analytics: A Technology Tutorial	Hu, H.; Wen, Y.; Chua, T.; Li, X.	2014	606	Big data analytics; Cloud computing; Data acquisition; Data storage; Data analytics; Hadoop
The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics	Ivanov, D.; Dolgui, A.; Sokolov, B.	2019	600	Supply chain dynamics; Supply chain risk management; Supply chain resilience; Supply chain design; Supply chain engineering; Industry 4.0; Additive manufacturing; Blockchain; Big data analytics; Ripple effect
Big Data: A Review	Sagiroglu, S.; Sinanc, D.	2013	568	Big data; Volume; Variety; Velocity; Verification; Value
Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations	Wang, Y.; Kung, L.; Byrd, T.	2018	556	Big data analytics; Big data analytics architecture; Big data analytics capabilities; Business value of information technology; Health care

Title	Authors	Year	Citations	Keywords
Machine learning in manufacturing: advantages, challenges, and applications	Wuest, T.; Weimer, D.; Irgens, C.; Thoben, K.	2016	537	Manufacturing; Machine learning; Intelligent manufacturing systems; Smart manufacturing
The future of manufacturing industry: a strategic roadmap toward Industry 4.0	Ghobakhloo, M.	2018	502	Information technology; Cybernetics; Strategic planning; Industry 4.0; Industrial robots
Big data and predictive analytics for supply chain and organizational performance	Gunasekaran, A.; et al.	2017	466	Big data; Assimilation; Routinization; Adoption; Supply chain performance; Firm performance
Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications	Hazen, B.; Boone, C.; Ezell, J.; Jones-Farmer, L.	2014	439	Data quality; Statistical process control; Knowledge-based view; Organizational information processing view; Systems theory
Viable supply chain model: integrating agility, resilience and sustainability perspectives-lessons from and thinking beyond the COVID-19 pandemic	Ivanov, D.	2020	422	Supply chain; Viability; Resilience; Sustainability; Lean; Agility; COVID-19; Pandemic; Recovery; Viable supply chain
A digital supply chain twin for managing the disruption risks and resilience in the era of Industry 4.0	Ivanov, D.; Dolgui, A.	2021	376	Supply chain; Resilience; Industry 4.0.; Disruption risk; Data analytics; Digital twin
Digital Supply Chain: Literature review and a proposed framework for future research	Buyukozkan, G.; Gocer, F.	2018	348	Digital Supply Chain; Literature review; Technology enablers
A review of Internet of Things (IoT) embedded sustainable supply chain for industry 4.0 requirements	Manavalan, E.; Jayakrishna, K.	2019	341	Sustainable supply chain; Closed loop supply chain; Enterprise Resource Planning; Industry 4.0.; IoT; IIoT

Title	Authors	Year	Citations	Keywords
How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management	Chen, D.; Preston, D.; Swink, M.	2015	335	Big data; Data analytics; Dynamics capability theory; Structural equation modeling; Survey research; TOE framework
Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management	Kache, F.; Seuring, S.	2017	334	Challenges; Supply chain management; Big Data Analytics; Delphi study; Opportunities; Digital information
Evolution of sustainability in supply chain management: A literature review	Rajeev, A.; Pati, R.; Padhi, S.; Govindan, K.	2017	314	Literature review; Sustainable supply chain management; Sustainable operations; Sustainability; Green supply chain management; Thematic analysis
The role of Big Data in explaining disaster resilience in supply chains for sustainability	Papadopoulos, T.; et al.	2017	314	Resilience; Big Data; Sustainability; Disaster; Exploratory factor analysis; Confirmatory factor analysis
Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications	Kamble, S.; Gunasekaran, A.; Gawankar, S.	2020	296	Agriculture supply chain; Food supply chain; Sustainability; Sustainable performance; Supply chain visibility; Big data; Blockchain; Data analytics; Supply chain resources
Blockchain applications in supply chains, transport and logistics: a systematic review of the literature	Pournader, M.; Shi, Y.; Seuring, S.; Koh, S.	2020	395	Blockchain; Supply chain; Logistics; Transport; Systematic review
Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives	Zhong, R.; Newman, S.; Huang, G.; Lan, S.	2016	283	Big Data; Service applications; Manufacturing sector; Supply Chain Management

Title	Authors	Year	Citations	Keywords
An Investigation of Visibility and Flexibility as Complements to Supply Chain Analytics: An Organizational Information Processing Theory Perspective	Srinivasan, R.; Swink, M.	2018	275	Analytics capability; Organizational flexibility; Supply chain visibility; Market volatility
Insights from hashtag #supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research	Chae, B.	2015	270	Supply chain management; Twitter; Data analytics; Network analytics; Content analytics; Big data; Social media analytics; Application Programming Interface
Big Data and Predictive Analytics and Manufacturing Performance: Integrating Institutional Theory, Resource-Based View and Big Data Culture	Dubey, R.; et al.	2019	265	Big Data; Predictive Analytics; Institutional Theory; Resource Based View; Big Data Culture; Manufacturing Performance;
Manufacturing and service supply chain resilience to the COVID-19 outbreak: Lessons learned from the automobile and airline industries	Belhadi, A.; et al.	2021	257	Covid-19; Supply chain resilience; Supply chain risk; Airline; Automobile; Financial impact
Smart factory performance and Industry 4.0	Buchi, G.; Cugno, M.; Castagnoli, R.	2020	255	Industry 4.0.; Fourth industrial revolution; Smart factory; Innovation; Value Chain; Enabling Technologies; Openness; Breadth; Depth; Performance; Opportunities; Regression models
Can big data and predictive analytics improve social and environmental sustainability?	Dubey, R.; et al.	2019	244	Big data; Predictive analytics; Dynamic capability view; Supply chains; Social sustainability; Environmental sustainability

Title	Authors	Year	Citations	Keywords
A critical investigation of Industry 4.0 in manufacturing: theoretical operationalisation framework	Fatorachian, H.; Kazemi, H.	2018	236	Industry 4.0.; Internet of Things; Cyber-physical systems; Manufacturing
Big data analytics in supply chain management between 2010 and 2016: Insights to industries	Tiwari, S.; Wee, H.; Daryanto, Y.	2018	233	Big data analytics; Supply chain management; Big data application
Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph	Tan, K.; Zhan, Y.; Ji, G.; Ye, F.; Chang, C.	2015	232	Big data; Analytic infrastructure; Competence set; Deduction graph; Supply chain innovation
Logistics 4.0: a systematic review towards a new logistics system	Winkelhaus, S.; Grosse, E.	2020	226	Logistics; Industry 4.0.; Smart; Internet of things; Big Data; Systematic literature review
Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential	Schoenherr, T.; Speier-Pero, C.	2015	227	Data science; Predictive analytics; Big data; Data scientist; Supply chain management; Education; Curriculum development
Supply chain risk management and artificial intelligence: state of the art and future research directions	Baryannis, G.; Validi, S.; Dani, S.; Antoniou, G.	2019	226	Supply chain risk management; Artificial intelligence; Decision-making; SCRM strategy; Supply chain disruption
Connecting circular economy and industry 4.0	Rajput, S.; Singh, S.	2019	219	Industry 4.0.; Circular economy; PCADEMATEL (Most likely); DEMATEL (pessimistic); DEMATEL (optimistic)
Management challenges in creating value from business analytics	Vidgen, R.; Shaw, S.; Grant, D.	2017	219	Analytics; Delphi; Management challenges; Value creation; Ecosystem

Title	Authors	Year	Citations	Keywords
Blockchain technology for enhancing swift-trust, collaboration and resilience within a humanitarian supply chain setting	Dubey, R.; Gunasekaran, A.; Bryde, D.; Dwivedi, Y.; Papadopoulos, T.	2020	216	Blockchain technology; Distributed ledger technology; Humanitarian supply chain management; Humanitarian operations management; Swift-trust; Collaboration; Supply chain resilience; Operational supply chain transparency
Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice	Arunachalam, D.; Kumar, N.; Kawalek, J.	2018	209	Supply chain management; Big data analytics; Capabilities;Maturity model
The impact of digital technologies on economic and environmental performance in the context of industry 4.0: A moderated mediation model	Li, Y.; Dai, J.; Cui, L.	2020	203	Digital technologies; Supply chain platforms; Industry 4.0.; Environmental dynamism; Economic performance; Environmental performance
Towards Industry 4.0 Mapping digital technologies for supply chain management-marketing integration	Ardito, L.; Petruzzelli, A.; Panniello, U.; Garavelli, A.	2019	202	Innovation; Marketing; Internet of Things; Patent analysis; Cloud computing; Supply chain management; Big Data analytics; Industry 4.0.; Cyber security; Supply chain management-marketing integration; Customer profiling
The performance effects of big data analytics and supply chain ambidexterity: The moderating effect of environmental dynamism	Wamba, S.; Dubey, R.; Gunasekaran, A.; Akter, S.	2020	200	Big data analytics; Dynamic capability; Supply chain ambidexterity; Supply chain agility; Supply chain adaptability; Organizational performance

Title	Authors	Year	Citations	Keywords
Big Data Analytics for Physical Internet-based intelligent manufacturing shop floors	Zhong, R.; Xu, C.; Chen, C.; Huang, G.	2017	200	Physical Internet; Big data; Radio frequency identification; Logistics; Intelligent shop floors
A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions	Ren, S.; Zhang, Y.; Liu, Y.; Sakao, T.; Huisingh, D.; Almeida, C.	2019	199	Big data analytics; Smart manufacturing; Servitization; Sustainable production; Conceptual framework; Product lifecycle
Big data applications in operations/supply-chain management: A literature review	Addo-Tenkorang, R.; Helo, P.	2016	189	Big data - applications and analysis; Internet of Things; Cloud computing; Master database management; Operations/supply-chain management
The applications of Industry 4.0 technologies in manufacturing context: a systematic literature review	Zheng, T.; Ardolino, M.; Bacchetti, A.; Perona, M.	2021	176	Industry 4.0.; Manufacturing systems; Advanced manufacturing technology; Manufacturing processes; Smart manufacturing; Literature review
Role of big data analytics in developing sustainable capabilities	Singh, S.; El-Kassar, A.	2019	175	Sustainable capabilities; Big data; Corporate commitment; Green supply chain management; Green human resource management practices; Environmental & organizational performance

The majority of the publications previously presented were entirely exploratory, giving readers an understanding of what Big Data Analytics is and what its primary uses are by analysing related Supply Chain Management technologies like Predictive Analytics, AI, and ML, while also reviewing a variety of other topics, including modeling of data-driven supply chains, sustainability, and supply chain resilience. These evaluations shed light on how Big Data might contribute to the revolution of traditional supply chains. Some of the chosen articles contributed significantly to this body of knowledge by reviewing major ideas that have been discussed in this review and by addressing the concerns of numerous experts in the field.

Multiple articles written by Angappa Gunasekaran and Thanos Papadopoulos offered some insights into how Big Data could assist organisations and supply chains in being more effective, efficient, data-driven, and resilient. These papers included various study evaluations and potential applications of Big Data in enhancing supply chain sustainability. Some of their most significant contributions include "*Big data analytics in logistics and supply chain management: Certain investigations for research and applications*" (2016), where the authors emphasise the significance of Big Data Analytics by analysing and categorising the literature on the use of Big Data in logistics and supply chain management and by stressing the need for managers to understand Big Data as strategic assets that should be integrated across business activities to enable integrated enterprise business analytics; and "*The role of Big Data in explaining disaster resilience in supply chains for sustainability*" (2017), where, using data collected from several social media websites, a theoretical framework is developed and put to the test to explain resilience in supply chain networks for sustainability, proving that information sharing, swift trust, and public-private partnerships are essential for enabling resilience in logistics systems.

Even though Dubey Rameshwar collaborated on several articles with the authors previously mentioned, his major contributions have more to do with how Big Data and Predictive Analytics interact to enhance supply chain performance, as these two technologies may tremendously benefit one another when used together, despite the scarcity of real-world applications on this subject. Research projects like "*Big data and predictive analytics for supply chain and organizational performance*" (2017), "*Big Data and Predictive Analytics and Manufacturing Performance: Integrating Institutional Theory, Resource-Based View and Big Data Culture*" (2019), and "*Can big data and predictive analytics improve social and environmental sustainability?*" (2019) help understand how top management commitment mediates the effects of resources (connectivity and information exchange) on organizational performance, Supply Chain Management, and Big Data assimilation (capability), raising attention to the value of big data and predictive analytics for better supply chain and operational performance, as well as the dearth of research on the impact of external institutional factors on an organisation's capacity to develop big data competence. In addition to examining the direct influence of Big Data Analytics in supply chain performance enhancement, Dubey also rests his research on understanding the impact of Big Data and Predictive Analytics on the organisation's social and environmental sustainability by empirically examining how these technologies affect social performance and environmental performance since very little is known about how they affect performance under specific contextual circumstances (such as business size and country environment and culture).

Numerous research on the implications of Big Data and digital technologies in Industry 4.0 was published by Dmitry Ivanov, making up a substantial contribution to this review's understanding of how supply networks may be reformed and re-planned in the future so that businesses can become more adaptable and resilient whenever big disruptions occur. Among all the contributions, "*The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics*" (2019) and "*Viable supply chain model: integrating agility, resilience and sustainability perspectives-lessons from and thinking beyond the COVID-19 pandemic*" (2020) stand out for emphasising how crucial it is to adopt Big Data along with emerging technologies like Predictive Analytics, Machine Learning, and Artificial Intelligence in order for managers to lower the associated risk when making decisions. These studies have gained even more significance in the last couple of years since a great deal of these research projects concentrate on the effects of adopting digital technology in risk management during the COVID-19 pandemic. As the first study to link business, information, engineering, and analytics perspectives on digitalisation and supply chain risks, Ivanov's research framework combines the findings from two previously independent areas, namely the impact of digitalisation on Supply Chain Management and the impact of SCM on the ripple effect control. These studies enabled the development of a "viable supply chain" model that decision-makers can use to create supply chains that can respond flexibly to both positive changes (i.e., the agility angle) and be able to absorb negative disturbances, recover, and survive during short-term disruptions and long-term, global shocks with societal and economic transformations (the resilience and sustainability angles). This approach, which emphasises that resilience is one of the critical elements in Supply Chain Management, can assist businesses in directing their decisions towards renovating and rebuilding their supply chains during major, protracted crises like the COVID-19 pandemic.

These publications, along with others that have not been outlined here, made it possible to construct a thorough literature review on the effects of incorporating Big Data Analytics in contemporary supply chains, helping researchers to comprehend how supply chain activities have advanced, contrasting old and new technologies, the significance of Big Data Analytics in current day-to-day business operations when combined with Artificial Intelligence and Machine Learning, and how gathering and handling information is an essential input step in these automated systems.

2.3 Findings

2.3.1 The advancement of Supply Chain activities

Over the last century, the global supply chain has changed dramatically. Every aspect of how businesses obtain, manufacture, and deliver products has been revolutionised [39]. From implementing new handling techniques to relying on long-distance deliveries and from containerisation to computerisation, global supply chains are more sophisticated and efficient than ever [40]. The introduction of the shipping container, and all the logistics and means of transport required to support it was arguably the greatest revolution in global supply chains, even if it would not be fully standardised until the late 1960s. The continuous innovation of pallets, handling equipment, containerisation, and other domains made freight transportation more efficient. Before 1960, logistical records and data were gathered, sent, and reported on paper [41]. Data computerisation began to streamline logistics, opening up potential in various sectors, such as more accurate forecasting, better warehouse storage, truck routing, and improved inventory management [42]. After 1980, technological developments, such as dynamic spreadsheets, mapping, and route planning, made it simpler to manage expenses and optimise revenues. This was accompanied by additional developments such as air freight optimisation, supply chain distribution networks, and the advent of Enterprise Resource Planning (ERP) systems [43].

The expansion of manufacturing in Asia has been one of the most recent influences on the restructuring of global supply chains, with China, Japan, and Korea becoming critical suppliers and exporters of commodities [39][44]. Combined with predictive analytics, Artificial Intelligence (AI) and Machine Learning (ML) provide businesses with better forecasting and order management processes, forcing industries to evolve toward a more data-driven, network-driven, and collaborative supply chain ecosystem that drives real value and growth for all participants [45][46][47][48]. With the development of infrastructure, globalisation, and technology came the introduction of Big Data (BD).

2.3.2 The usefulness of Big Data Analytics

Big Data Analytics (BDA) implies two perspectives: big data and business analytics [49][50]. Big data refers to high-volume, high-velocity, and high-variety amounts of dynamic data that surpass the processing capabilities of standard data management systems [51]. Business analytics is the study of the skills, technologies, and procedures used to continually analyse organisation-wide plans and operations to gain insights and steer an organisation's business strategy [19][52][53]. Through evidence-based data, statistical and operational analysis, predictive modelling, forecasting, and optimisation methodologies, such a review might extend from strategic management to product creation to customer service improvement [54][55][56][57]. Predictive Analytics (PA) is a subset of analytics that uses input data and statistical combinations to estimate the likelihood of a specific event occurring, anticipate future trends or outcomes using on-hand data, and ultimately enhance a corporation's performance [13][58][59][60]. PA technique enables organisations to be proactive, forward-thinking, and anticipate outputs and behaviours based on facts rather than assumptions with no supporting data or information [61][62][63].

According to Tobias Schoenherr, and Cheri Speier-Pero, Big Data, and Predictive Analytics may help reduce supply chain costs and achieve better levels of efficiency, adapt faster to changing environments, provide more leverage in supplier relationships, and improve sales and operations planning skills [62]. However, due to the continuous development of these activities in SCM environments and the fact that the data these functions rely on is frequently riddled with mistakes, there is also an urgent need to investigate the many issues regarding data quality [64][65].

Avita Katal, Mohammad Wazid, and R. H. Goudar defined Big Data as having four distinct characteristics (4 Vs): Volume, Velocity, Variety, and Value [66]. Variety suggests that BD contains a wide range of data types and that this variety splits the data into structured and unstructured data [67]. Volume alludes to the enormous amount of data [68]. Fosso-Wamba et al. defined volume as the large amount of data that either consumes huge storage or entails of a large number of records data [69]. Large-scale data in SCM may refer to, for example, data from Radio frequency identification (RFID) transmitters and other types of sensors used for product identification and transportation, mobile phone Global Positioning System (GPS) signals, and purchase transaction records [70][71]. Velocity refers to the rate at which data is generated and the speed at which it should be analysed and acted upon [72]. Value denotes usefulness and is inversely related to the overall size of the data set: the more significant the size, the less valuable the data is [73]. Firms should recognise the vast quantity of data available and determine what is helpful to extract for further research. A study revealed that Tesco boosted its operational margins by analysing Big Data connected to temperature and weather trends, allowing them to undertake better temperature forecasts and corresponding changes in consumer demand [74].

By applying such analytics to massive datasets, important information may be collected and used to improve decision-making and support informed judgments. As the volume of Big Data expands tremendously, companies across all industries are increasingly interested in knowing how to handle and analyse such data [75]. As a result, businesses are rushing to capitalise BD's opportunities to acquire the most advantage and insight possible. As a result, they are adopting Big Data Analytics to unlock economic value and make better and quicker choices [73][76]. In due course, firms are turning to BDA to analyse massive volumes of data quicker and uncover previously unknown patterns, attitudes, and consumer information [77]. Scholars have paid close attention to BD and SCM, but BD research is still in its early stages, and there is an urgent need for a study to distinguish high-quality data sets from low-quality data sets [39][78].

There are considerable gaps in the literature, particularly with Machine Learning and Blockchain approaches for Supply Chain Management applications (most articles often miss a supporting case study). There is also a scarcity of academic work that presents the future advantages and conceptualisations of Big Data's role in sustainability and resilience. More significantly, no studies use Big Data (unstructured data) to explain disaster resilience for sustainability.

2.3.3 Understanding the Extract, Transform and Load process

Most businesses have a large amount of data stored in various systems such as ERP or Customer Relationship Management (CRM) software. This amount of data is rapidly expanding due to increased digitisation [79]. Since raw data is never suitable for analytics or Machine Learning processes, it must be transformed into a usable shape and form. Companies often use an Extract, Transform and Load technique to solve this issue, which is beneficial for processing massive data volumes [80][81]. ETL is an acronym for Extract, Transform, and Load. The goal is to prepare and provide the combined data for subsequent processing. This procedure is applicable when information is scattered across several subsystems, is redundant, or has a distinct structure. The heterogeneously organised data from many sources is integrated and prepared throughout the process. In the data warehouse, data quality is assured, and consistency is established [82].

This course of action is separated into three stages: during the extract phase, data is taken from the source system and placed in the staging area [83]. Transformations, if any, are performed in the staging area to ensure that the performance of the source system is not compromised. During the transformation phase, a sequence of rules and functions are applied to the extracted data to prepare it for loading into the end target [80]. Data cleansing, which tries to transmit error-free data to the destination, is a crucial transformation function. The data is loaded into the final target during the load phase. This approach varies greatly depending on the company's needs [81]. Some data warehouses may replace current data with cumulative data, meaning that extracted data is often updated daily, weekly, or monthly. Other data warehouses may contribute new historical data at regular intervals, such as hourly or daily.

2.3.4 The adoption of Machine Learning and Artificial Intelligence Techniques

It is commonly understood that collecting high-quality data from these vast, varied, and complex data sets has become a crucial process confronted with several obstacles, such as poor data integration, short longevity and usefulness of the collected information and limited existence of unified data quality standards [84]. Business partners are now using Artificial Intelligence approaches to address some of the difficulties associated with the large volume of existing data. AI contributes to data velocity by enabling quick computer-based judgements that lead to subsequent decisions [85][86]. AI researchers have long been interested in developing apps that analyse unstructured data and categorise or arrange that data in some way so that the resulting insights may be utilised directly to understand a particular process. Machine Learning algorithms are frequently used in these applications [87][88]. ML enables users to uncover the underlying structure and make predictions from large data sets [89]. ML thrives on efficient learning algorithms, massive data, and powerful computing environments. As a result, ML has enormous potential and is a vital component of Big Data Analytics [90].

Lina Zhou et al. introduced a framework of Machine Learning in Big Data contexts [87]. The ML component is at the heart of the framework and interacts with four other components: Big Data, user, domain, and system [91]. Interactions occur in both ways. For example, BD is being fed into ML, and the latter generates outputs that become part of Big Data; users can interact with Machine Learning by providing domain knowledge, personal preferences, and usability feedback, as well as leveraging learning outcomes to improve decision-making; domain can serve as both a source of knowledge to guide ML and the context in which learned models are applied; system architecture influences how learning algorithms should run and how effective they are.

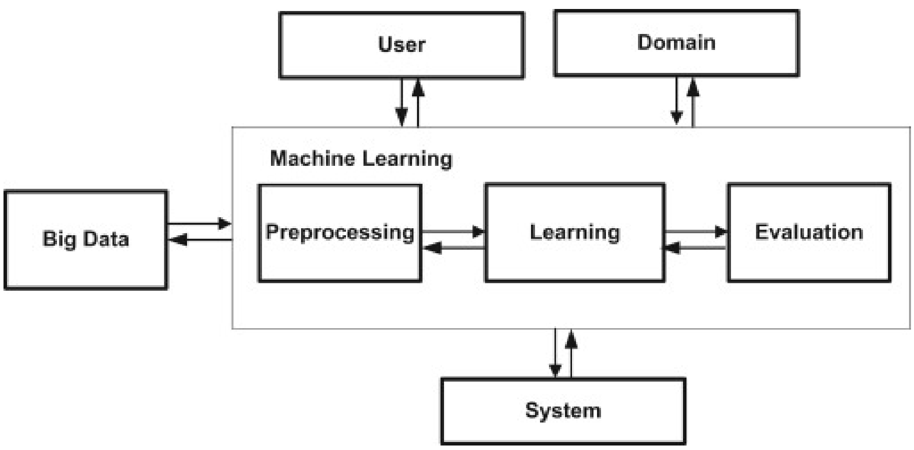


Figure 2.6: The framework of machine learning on Big Data (based on Zhou, Pan, Wang & Vasilakos, 2017).

Stakeholders in Machine Learning systems include domain experts, end users, researchers, and practitioners. Traditionally, ML practitioners make most Machine Learning-related choices, from data collection to performance evaluation. End-user participation in this process has been restricted to supplying data labels, answering domain-related questions, or providing feedback on learned findings, often mediated by practitioners, resulting in lengthy and asynchronous iterations. Involving all users in ML can result in more effective learning systems and better user experiences. For example, interactive Machine Learning allows users to analyse the impact of their actions in real-time and alter subsequent inputs to influence ML behaviours to achieve desired results [92].

Domain knowledge assists Machine Learning in uncovering unique patterns that would not be discovered with only datasets. Obtaining appropriate and representative data is costly, if not impossible, due to domain variance and application-specific needs. Domain knowledge can aid in the generality and robustness of patterns derived from datasets. A system design generates an environment in which ML algorithms may execute. To solve the issues of extensive data, new frameworks and system architectures such as Hadoop or Spark have been developed [93]. However, moving old ML algorithms to distributed architecture necessitates changes in how these algorithms are designed and delivered. Furthermore, the distinct goals and values of Machine Learning may motivate the design and development of innovative system architecture [94][95].

2.3.5 Understanding the concerns regarding data quality

The expected expansion of AI solutions is a source of concern for Supply Chains [96]. Algorithms, at their root, use data to create predictions and modify models as new data is obtained. AI models can have a bias due to developer design, and the model continues to learn and adapt with minimum interaction once deployed [97]. Because a model cannot discriminate the accuracy of data presented, generated forecasts and alternatives may be incorrect, and there is no mechanism for Supply Chain experts (or anyone else) to track back through iterations to discover flaws [98][99]. When a defect is found at a distribution center, incorrect logistics data may result in minimal additional expenses. However, when the supplies reach the customer's location, this data flaw may result in significant extra expenditures (for the return of the goods and lowered service level) [100], suggesting that the kind of data error and the moment it is discovered appear to influence its commercial impact. Modelling and monitoring a data life cycle perspective of a company's ecosystem, which includes all create, read, update, deactivate, and archive activities, would allow for model-based analysis of data flows and data maintenance processes and aid in identifying weak points in data maintenance processes.

Big Data paired with Predictive Analytics has shown to be a robust advancement that will influence supply chain structure and management practices [101]. Furthermore, as additive manufacturing (also known as 3D printing) renders traditional production, distribution, and demand models obsolete in various product categories, understanding the uses and repercussions of BD and PA will become increasingly crucial [102][103][104]. To address some of the challenges related to collecting high-quality data, Li Cai and Yangyong Zhu proposed a data quality framework to formulate a dynamic Big Data quality assessment process with a feedback mechanism [96] (Figure 2.7).

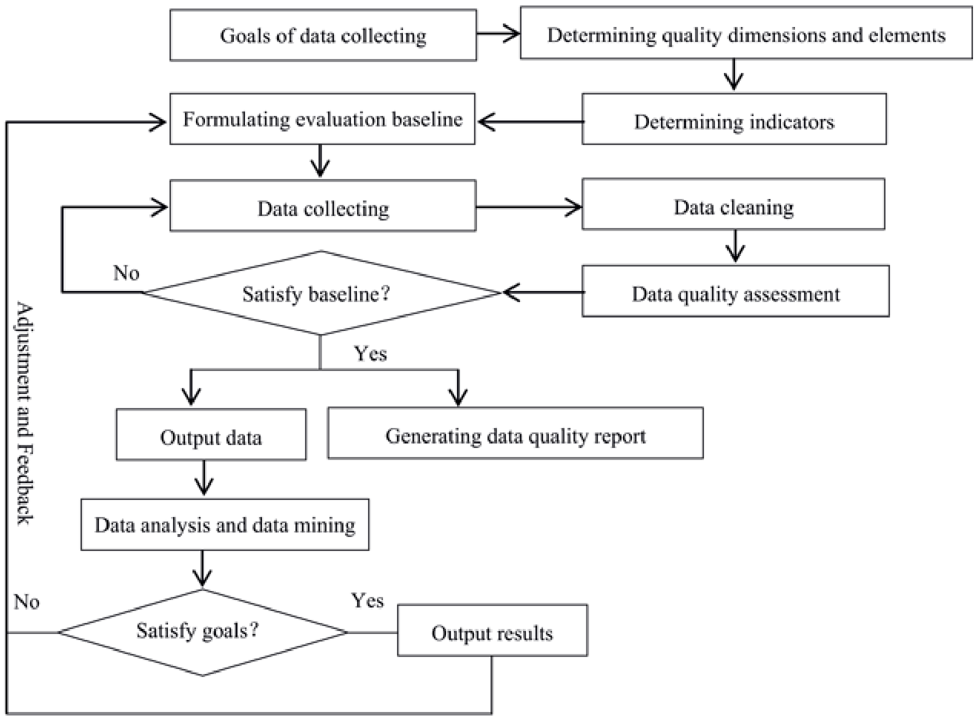


Figure 2.7: The quality assessment process for Big Data (based on Cai & Zhu, 2015).

The first stage in the evaluation process is determining the data collection goals. Data sources, type of data, volume, quality requirements, assessment criteria, and specifications should all be specified in advance. The formulation of assessment indicators depends on the actual business environment. These require the data to comply with specific conditions or features. The preliminary assessment findings of data quality dimensions set the baseline, while the remaining evaluation is utilised for continual detection and information enhancement as part of the business process. After the quality assessment preparation step, the procedure moves on to the data gathering phase. However, most data obtained are not always helpful (data errors, missing information, inconsistencies, noise) [105]. To address this, a data cleaning operation is used to find and eliminate flaws and inconsistencies in data to enhance its quality. This procedure consists of several steps, including normalising capitalisation, converting data types, and clearing excessive formatting [106].

The workflow then moves on to the data quality evaluation and monitoring phases. How to analyse each dimension is crucial to data quality evaluation. The model approach is divided into two categories: qualitative and quantitative. From the standpoint of qualitative analysis, the qualitative evaluation technique is based on specific criteria and requirements according to assessment objectives and user requests. The quantitative approach is a formal, objective, and systematic process for obtaining knowledge using numerical data. Following evaluation, the data may be compared to the previously defined baseline for data quality assessment. If the data quality meets the baseline criterion, the phase of follow-up data analysis can begin, and a data quality report will be created. Otherwise, new data must be obtained if the data quality does not meet the baseline criteria.

2.3.6 The impact of technology on Supply Chain processes

It has been suggested that competition is no longer between enterprises but within the whole supply network. As a result of the increased focus on Supply Chain Management, managers are being compelled to rethink their competitive strategy [107]. Because both technology and data are available, businesses must choose how to use both to win. Data is becoming increasingly important to Supply Chain managers for obtaining visibility into expenditure, recognising patterns in costs and performance, and enabling process control, inventory monitoring, production optimisation, and process improvement activities [108]. Big Data improves visibility by offering an integrated framework for monitoring performance and customer contact via real-time data analysis and critical decision-making scenarios, decreasing risk and Supply Chain interruption and failures [14][109][110].

Big Data technology-enabled operational excellence may boost productivity and assist organisations in the Supply Chain industry to gain a competitive advantage [111][112][113]. Analytics can improve customer satisfaction and, in the long run, help retain consumers [114][115][102]. The Internet of Things (IoT) gives real-time visibility, mirrored more broadly throughout the industry [116]. Better vendor selection methods combined with qualitative and quantitative performance analysis utilising BD technologies might provide cost reductions and low-cost operations. Big Data may also assist in closing the gap between the demand and Supply Chains by providing data-backed demand projections and insights into client purchasing behaviour [117].

Data quality assurance techniques have grown dramatically in the last decade. Assessing data quality can be a good starting point for identifying insignificant information. Data quality is often characterised as fitness for use, or a data collection capacity to fulfil users' needs [118]. This fulfilment is evaluated by a limited selection of data quality parameters considered necessary in most modern research. This set includes accuracy, completeness, timeliness, and consistency [119].

Assessment techniques for parallel computing and minimising computation must be redesigned to overcome Big Data volume and velocity challenges. As a result, Danilo Ardagna et al. suggested limiting the computing area by considering only a subset of the data as input, reducing the time and resources required [120]. A confidence metric was developed to measure data trustworthiness to inform consumers of the integrity of the Data Quality values. Confidence is proportional to the amount of considered data, determined by the execution time limitations controlled by the available computational resources.

The researchers utilised the EUBra-BIGSEA project as an illustrative case study [121]. This project aims to provide a cloud platform that promotes creating and managing Big Data applications. The objective is to offer cloud services that enable Big Data Analytics and aid in developing data processing applications [122]. Such services are created by considering the previously mentioned BD concerns (data volume, variety, velocity, and veracity), quality of service, privacy, and security restrictions [123]. Within the context of the previous project, a solution was developed in response to these difficulties. This type of solution is called Data Quality service and primarily provides evaluation capabilities (Figure 2.8).

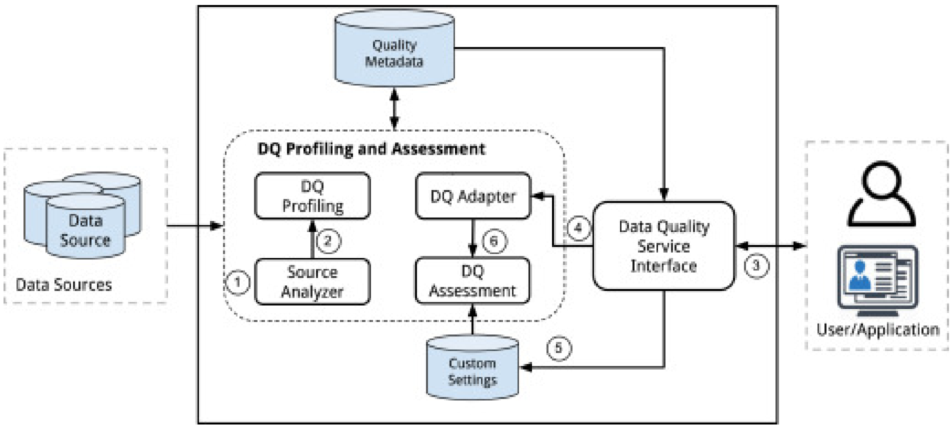


Figure 2.8: Data quality service architecture (based on Ardagna, Cappiello, Samá & Vitali, 2018).

The data quality profiling and assessment module lies at the heart of the design. This module is in charge of gathering data, statistics, and information. Profiling is responsible for producing a collection of metadata that specifies the data source, the significant properties of the fields that comprise it (e.g., maximum, minimum, average values), and the number of distinct values. The assessment module computes Data Quality dimensions. The metrics are generated based on the data component chosen for the individual application and the quality metrics relevant to the end user. Because the nature of the data source and the type of its properties vary, not all quality metrics are computable given a data source or a piece of it. After implementation, the architecture can assess the previously specified quality dimensions, such as:

- **Accuracy**: The degree to which a value is correct is defined. The authors have only implemented it for numerical values during their case study to determine if they are within a scheduled interval or are outliers (and hence inaccurate).
- **Completeness**: Determines the extent to which a data set is complete. It is calculated by comparing the number of values accessible in the data set to the predicted number. It should be noted that missing registrations in data streams are easier to identify if data is detected at a specified frequency.
- **Timeliness**: the extent to which values are valid at the moment. It is calculated considering the last update timestamp and the data's average validity.
- **Consistency**: The breach of semantic norms specified over a group of data elements. As a result, this dimension can only be computed if a set of rules representing attribute dependencies is available. The architecture provided already contains a module that discovers functional relationships and examines whether the values in the data set respect them.

The studies indicated that the sensitivity of the data quality evaluation to confidence relies on both the data source attributes and the individual data quality dimension; various dimensions might be responsive to the confidence level in different ways.

Organisations' operations and behaviours will progressively shift due to digital transformation [124][125]. Companies must embrace automation, Artificial Intelligence, and cloud operations to maximise the transformation's influence [126]. The practical application of these technologies will demand a significant change in company structures, attitudes, and work cultures [127][128]. Data protection rules are still in their inception. Regulators are still working to address the ethical issues that Big Data raises and numerous (and developing) technologies emerging to extract and exploit its meaning. This increases the strain on data-driven organisations' capacity to manage and govern data [129]. The regulatory landscape has evolved to the point where avoiding age-related data storage errors is required rather than corrective action [130].

Chapter 3

Interviews

3.1 Research Methods

Different research approaches can be distinguished based on their topology [131]. For instance, when it comes to data collection and analysis, quantitative research involves working with amounts and percentages, whereas qualitative research is concerned with words and meanings [132]. Both of these approaches are complimentary methodologies that may be used in surveys to get broad and deep findings [133].

Quantitative research is expressed in statistics and graphs [134][135]. It is used to test or confirm theories and assumptions. This research method can be used to establish evident facts about a subject. Common quantitative approaches include experiments, observations recorded as numbers, and surveys with closed-ended questions. Information bias, omitted variable bias, sampling bias, and selection bias are all risks in quantitative research [136][137].

Qualitative research is expressed in words. It is used to understand concepts, thoughts or experiences [138]. This form of study allows you to gain in-depth knowledge about issues that are not generally understood. Interviews with open-ended inquiries, and observations documented through words are common qualitative approaches [139]. Certain study biases, such as the Hawthorne effect, observer bias, recollection bias, and social desirability bias, can also occur in qualitative research [140].

As pointed out in Chapter 2, the study research revealed significant gaps in the literature, particularly regarding Machine Learning and Blockchain techniques for Supply Chain Management applications. Many evaluated publications lacked a single supporting case study, implying that every component of the debate was solely theoretical and had no means of being validated at the time.

Since the primary objective of this dissertation is to enlighten the scientific community with a current literature review of the application of Big Data Analytics in SCM methods, providing knowledge from previous articles, and proposing future research cases for different industries, it was decided to conduct numerous interviews with multiple field experts in order to support (or contradict) the findings discovered throughout the literature research procedures. These interviews were conducted with eighteen Logistics and Supply Chain Management professionals, such as teachers, researchers, factory managers, and several operations directors, working in various industries, including retail, transportation and logistics, automotive, and engineering (technology). Due to privacy concerns, all the names of the professionals questioned were kept undisclosed, since many insights presented could be helpful for the companies' competition. Even though 18 professionals is a small sample compared to the number of scientific researchers working on these topics, it is necessary to examine the various points of view and perspectives regarding using Big Data in the many industries presented.

3.2 Bardin's Content Analysis

In-depth interviews are a qualitative research technique designed to allow interviewees to share their narratives in their own words. Conversely, the interview is dedicated to a particular topic and seeks to inform listeners by answering predetermined questions. Multiple authors advise that the researcher should not prepare a complete list of questions but rather have an understanding of the domains in which the interviewee has more expertise and feels more comfortable discussing openly, as well as be able to tie this knowledge to the ongoing study [141]. As a result, it is required to write an organised script. This guide might be precise, with carefully constructed questions, or it can be a list of subjects to address. Either way, the topics should always be tied to the significant issue explored in both situations. Given the various industries, a script with several research questions was prepared for this particular scenario to acquire an overview of the different points of view on critical subjects.

In this study, a content analysis technique was used. Content analysis, as defined by Bardin (1977), is a set of communication analysis techniques aimed at obtaining, through systematic procedures and objective description of the content of the messages, indicators (quantitative or not) that allow the inference of knowledge related to the conditions of production/reception (inferred variables) of these messages [142]. According to Bardin, this analysis foresees its use in three distinct stages:

1. **Pre-analysis:** a broad review of all the information acquired for analysis is performed, and in the case of the interviews, this consists of transcript analysis. It employs procedures such as document selection, hypothesis and objectives formulation, and creation of key indicators to substantiate the final interpretation;

2. **Material exploration:** identification of reporting units, context units, and challenges raised by the readings, as well as converting text samples collected from transcripts into recording units for future study and categorising the acquired material;

3. **Treatment of results and interpretations:** identify deductions and develop interpretation concerning the specified aims, or which relate to other unanticipated findings.

3.3 Interview Questions

The semi-structured script opens with an introduction that informs the interviewee of the investigation's subject and its main objectives, as well as asks for his consent to record his answers and permission to take notes while the interview is taking place. The script's questions should center on the major research topics, such as "What is the true impact of using Big Data Analytics on modern companies?" and "What are the benefits and drawbacks of using Big Data Analytics derived processes?". All of the follow-up questions should encourage the interviewee to elaborate on specific topics, such as supply chain digitisation, the use of Artificial Intelligence and Machine Learning on logistical processes, the difficulties of information sharing, and the application of Blockchain technology further to bolster interactions between businesses in terms of security. Table 3.1 provides a more comprehensive overview of all the prepared questions:

Table 3.1: Interview questions.

Question	**Reasoning**
1. What role do you think Big Data can have in modern supply chains? What are the major benefits and drawbacks of utilising this data for business activities?	Intended to get the interviewee talking about the broad uses of Big Data in modern daily operations, given multiple industries, as well as to gain a deeper understanding of the benefits and difficulties associated with the implementation of these techniques.
2. What are your thoughts on information sharing between the various supply chain stakeholders, given the significant paradigm change between traditional and present-day supply chains and the fact that many businesses take a competitive and non-collaborative stance?	Attempts to explore the influence of stakeholder collaboration on modern supply chains, as in the past, the majority of industry sectors often maintained a competitive position against their competitors, operating alone with the purpose of maximising their own profitability.
3. To what degree can the use of technologies such as AI and ML assist businesses in making more assertive decisions? What are the main benefits and difficulties that arise from the adoption of these technologies?	It is important to understand the influence of Artificial Intelligence and Machine Learning on Big Data Analytics operations. These technologies may be utilised to automate the majority of the data collection and treatment steps, providing enterprises with an almost ideal bundle of usable data to work on.
4. How crucial is it to establish a more transparent and environmentally friendly supply chain given the intense pressure from stakeholders nowadays? Could Big Data contribute in the supply chain's decarbonisation?	Attempts to highlight the importance of decarbonisation and making modern supply chains more environmentally friendly, as several governments and shoppers are pressuring businesses to decrease their ecological footprint.
5. What benefits do information systems provide for businesses? Is there a significant effect on how fast companies grow?	Intended to comprehend the effects of information systems on product transportation and/or warehouse management procedures, as multiple industries demand varying degrees of information.
6. What new technology or technique may soon cause a disruption in the present supply chains, abolishing certain conventional processes while introducing completely new requirements and paradigms?	Seeking the interviewee's perspective on the next major technological advancement that will significantly alter the traditional supply chain, either by influencing manufacturing procedures or major supply chain logistical activities.
7. From your experience, what are the main pain points in modern supply chains? What role may the processing of analytical data and the usage of BD play in resolving these issues?	Designed to obtain the interviewee's concluding thoughts on the primary uses of Big Data Analytics to address specific supply chain pain points.

3.4 Study validity - the concept of saturation

Data saturation refers to the stage of a research process when sufficient data have been gathered to reach the appropriate conclusions and further data collecting will not result in new, insightful information [143]. The phrase "data saturation" comes from the qualitative research approach known as "grounded theory," which was initially introduced by sociologists Glaser and Strauss in the 1960s [144]. In other words, no matter how comprehensive the study is, the same themes will emerge when the quantity of interviews (in this case) is sufficient. In contrast to quantitative research, where preexisting hypotheses serve as a framework for research, the grounded theory describes how patterns in data from social research can be used to generate theories and hypotheses on which additional research can be conducted [145].

In 2006, Guest, Bunce, and Johnson finally addressed the concept of saturation in qualitative research studies by providing a precise, straightforward, repeatable, and reliable way to do so in an essay that has since proven essential to qualitative research [146]. They created an approach defining saturation as a ratio, or the number of categories found at a particular study stage divided by the total number of categories found across the whole sample. When 80–90% of the categories in a data set have been recognised, the saturation level is said to have been reached. They calculated the number of new categories per set of six interviews based on a qualitative study conducted in two African nations (Ghana and Nigeria). They discovered that this value declines as more interviews are undertaken. However, what is more surprising is the speed at which this decrease in yield occurs. New codes become less common after the 18th interview and practically vanish after the 36th interview. According to Guest et al. (2006), saturation can occur after the sixth interview, which increases the likelihood that researchers can choose the appropriate sample size early in the research development process [146]. They also concluded that all datasets would always be able to attain 100% of the saturation since the total number of interviews is a fixed denominator, and the numerator is inevitably getting closer to the denominator with each new interview (according to the formula explained previously) [147].

As previously explained, 18 participants were contacted to inquire about the implementation and applications of Big Data Analytics within modern supply chains. The general information about each participant, including his job title and business sector, is displayed in Table 3.2:

Table 3.2: List of participants.

Participant	Occupation	Industry
A	Global Procurement Operations Director	Retail
B	Global Senior Head of Procurement	Retail
C	Procurement Manager	Retail
D	Logistics Director	Retail
E	Chief Operating Officer (C.O.O.)	Retail
F	Chief Operating Officer (C.O.O.)	Transport and Logistics
G	Chief Executive Officer (C.E.O.)	Transport and Logistics
H	Innovation and Information Systems Director	Transport and Logistics
I	Business and Logistics Director	Transport and Logistics
J	General Director (Country)	Logistics Service Provider
K	Chief Operating Officer (C.O.O.)	Automotive
L	Chief Operating Officer (C.O.O.)	Automotive
M	Logistics Director	Automotive
N	Global IT Manager	Information Technology (IT)
O	Managing Partner	Information systems (Software)
P	Vice President of Customer Success	Information systems (Software)
Q	Teacher / Researcher	—
R	Teacher / Researcher	—

In order to categorise all of the interviews using the coding approach suggested by Bardin, a categorical analysis was carried out to address the primary issues and objectives that this study proposes [142]. Following the selection of the information and a cursory review, the categories were created. This initial phase of code building to create analysis categories is carried out by a method in which the researcher emphasises information from the obtained data. This enables the researcher to recognise characteristics, dimensions, and ideas analytically across all transcriptions. Since most thoughts could not be solely broken down into individual words, most encodings took the form of snippets, sentences, and/or concepts, as shown in Table 3.3. Applications like NVivo and VOSViewer (previously used in bibliometric analysis to construct linked networks) were used to carry out this encoding.

Table 3.3: Interview categories.

Category	**Amount of interviews**	**% of total interviews**
Big Data enables businesses to take more decisive actions	18	100%
Information exchanged among multiple stakeholders adds value for businesses	18	100%
Supply chain operations will be disrupted by Artificial Intelligence and Machine Learning	18	100%
The future lies in automation	18	100%
Big Data Analytics must be an automated digital process	18	100%
Digital information systems are required in order to implement an agile approach since manual ones would require employees to manually verify the amount of goods in storage	17	94%
Prior to implementing BDA, companies must identify their key objectives	16	89%
Companies with less financial clout must be aware of the expenses associated with adopting BD strategies	16	89%
Businesses can estimate demand and eliminate uncertainty with the help of Big Data Analytics	16	89%
Information systems assist businesses in keeping track of their spending and stock levels	16	89%

Category	Amount of interviews	% of total interviews
Each operation carried out within a business should provide data for subsequent studies	16	89%
Data uniformization is a crucial step in data exchange amongst stakeholders	15	83%
Companies collect a lot of information, both insightful and meaningless	15	83%
Through digitalisation, supply chains will become more transparent	15	83%
If businesses lack the capacity to acquire vast volumes of data, they can review information that is freely accessible	14	78%
As time passes, information loses value	14	78%
Big Data could undoubtedly assist businesses in being more ecologically conscious	14	78%
If businesses collaborated more, implementing Big Data could become simpler	14	78%
Configuring the information treatment algorithm appropriately is one of the most difficult aspects of Big Data implementation	13	72%
Information may help consumers and producers develop tighter ties	13	72%
Different sectors require different degrees of information	13	72%
Big Data inexperience is an obstacle for companies	12	67%
Businesses are compelled to use an agile approach by Big Data	12	67%
Autonomous systems will provide managers with fewer best-case scenarios to choose from, allowing for more efficient decision-making	12	67%
Businesses are going to deliver better services as a result of using Artificial Intelligence and Machine Learning since they are less likely to make mistakes	11	61%
Big Data Analytics and Artificial Intelligence may be able to assist businesses identify buying trends	11	61%

Category	Amount of interviews	% of total interviews
Some supply chains don't necessarily require a high level of information exchange since they do not employ a just-in-time methodology	10	56%
By consolidating information systems, organisations may operate more cost-effectively and with fewer disruptions	10	56%
Big Data eliminates the need for middlemen in the transport of goods	10	56%
The automobile sector is at the forefront of Big Data Analytics applications	10	56%
Due to inefficient scheduling and delays in cargo transit and exchanges, goods are frequently stranded in terminals	10	56%
Automation would enable businesses to be more environmentally friendly	10	56%
Portuguese harbours lack automation as a result of previous government under investment	9	50%
Blockchain technology can assist organisations in sharing information more securely	9	50%
Using Big Data to decrease risk and uncertainty will assist organisations in lowering their prediction percentile variation	9	50%
Organisations are unaware of the value of Big Data	8	44%
Similar visibility standards for B2B and B2C transactions should be implemented	7	39%
Social media enables businesses to focus on their target audience	7	39%
Businesses may assess the quality of their services using social media analytics	6	33%
Companies typically store sensitive information offline out of concern for cyberattacks	6	33%
Companies must have a contingency strategy before implementing data-derived approaches	6	33%

Category	Amount of interviews	% of total interviews
Even with Big Data Analytics methods in place, companies often experience uncertainties about the data gathered	6	33%
Big Data implementation is impossible when there are too many information systems in use	6	33%
Employees of automotive businesses may benefit from ML in assembly line processes	5	28%
In a perfect world, new supply chains should be proposed rather than existing ones being modernised	5	28%
Automation would allow organisations to hire more personnel for administrative positions	5	28%
Government regulations (collisions, dangerous to humans) will prevent drone massification from being successful	4	22%
Humanoids within warehouses are an excellent alternative to completely automated storage facilities since they are less expensive and more manageable	4	22%
Artificial Intelligence algorithms might automate rail transit, removing the need for a driver	3	17%
In the future, it will be crucial to eliminate paper from assembly line procedures	2	11%

As can be seen from the preceding tables, the transcript analysis of the interviews resulted in the collection of several thoughts and ideas, regardless of being generalised or exclusive to one industry. Given this, an assessment was performed to address the concept of saturation and the discovery of new categories in this sort of qualitative research, as Guest, Bunce, and Johnson demonstrated earlier. A base size of four interviews and a run duration of two interviews were used to calculate prospective data saturation. For this demonstration, a new information threshold of 5% was chosen to denote that the saturation level had been met. In addition to showing the base, runs, and saturation points, Figure 3.2 includes the data used for each step.

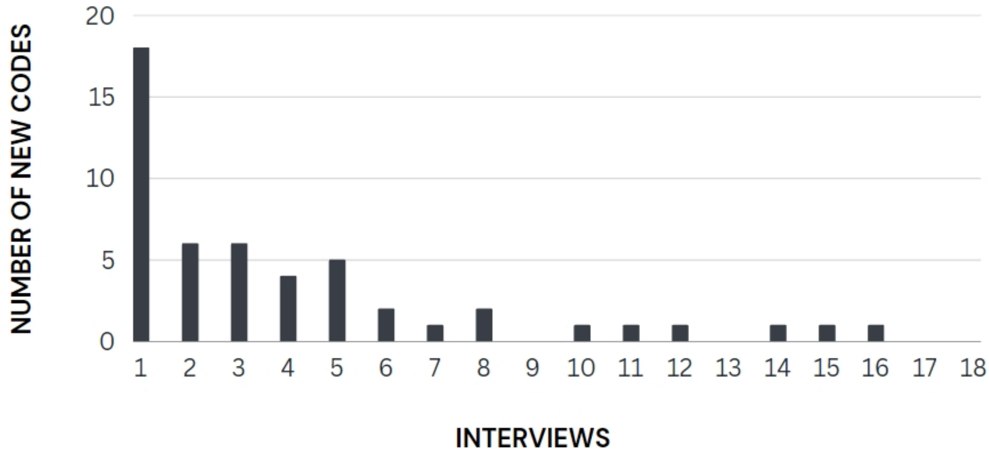


Figure 3.1: Amount of new categories per interview.

As seen in Figure 3.1, 34 unique themes have been found in the first four interviews, representing the base number of our analysis. Since we are using a run length of two interviews, we can then calculate the number of new codes per run of two conversations, allowing us to determine the proportion of new codes found over the base denominator (34). After doing identical calculations throughout all interviews, it is discovered that after the 10th interview, the percentage of new data added by the previous run is less than the predetermined threshold of 5%, having a solid feeling that the amount of new information is dwindling to the point where, according to our subjective indicator, saturation has been attained.

Interview number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
New codes per interview	18	6	6	4	5	2	1	2	0	1	1	1	0	1	1	1	0	0
New codes per run of interviews				34		7	3	3	2	1	2	2	1	1	2	2	1	0
% change over base						21%	9%	9%	6%	3%	6%	6%	3%	3%	6%	6%	3%	0%

Figure 3.2: Prospective data saturation calculation.

These illustrations show that after the first interview, the number of new categories reduces significantly and stays at a reasonably low value after the fifth interview. After the tenth interview, the saturation level was attained, supporting the conclusions of Guest, Bunce, and, Johnson, as 45 of the 50 categories were discovered (90%). This does not imply that interviews conducted after the tenth one were pointless or ready to be neglected, but rather that several significant ideas were repeated (as shown in Table 3.3) regardless of the experts' main industries.

3.5 Key takeaways

All participants agreed that Big Data Analytics is a crucial tool for business growth since it allows organisations to examine all data related to operations like manufacturing, storage, and delivery, providing them with more knowledge about what can be optimised and what issues still need to be solved. The guest speaker "J" highlighted the importance of Big Data in present-day supply chains by saying that "*if companies implement Big Data Analytics methodologies and have the information properly processed, structured, and with clear strategic objectives, the information will be the fuel to reach these goals, like anticipating events, anticipating trends, mitigating risks, reducing lead-times, reducing losses, and reducing costs.*" In addition, interviewee "Q" noted that the retail and automotive industries account for the majority of Big Data Analytics applications in Portugal, since these industries frequently employ Just in Time methodologies, which require quick and efficient work from every network member.

Several respondents agreed that Big Data Analytics is crucial when directors must make strategic decisions, with some saying that "*Big Data has gained a new dimension, as it is crucial to understand the consequences of taking certain decisions, since assessments are frequently made based on gut or experience*". In the present market, where there is more rivalry among businesses, it is vital to rely more on the procedures than on the decision-makers themselves. However, it is also critical for business managers to comprehend that Big Data strategies must be well-considered and adequately implemented, given that it costs a lot of money and effort for companies to develop, deploy, and maintain these technologies, with participant "M" claiming that "*the major challenges in adopting Big Data are connected to the expenses of implementation and the lack of in-depth understanding of the processes*". Participant "A" had an identical perspective on the difficulties of using Big Data without enough planning, noting that "*businesses frequently overlook the groundwork when putting together Big Data Analytics solutions, concentrating primarily on the output rather than the input data*".

While Big Data Analytics may be a valuable strategy for businesses, collecting and processing the necessary data to feed these processes and obtain their output is essential. This is where business-to-business information sharing becomes vital since it enables multiple parties operating within the same supply chain network to collaborate and work on the same information simultaneously, giving organisations a high degree of visibility and transparency, with interviewee "L" highlighting that "*in order to be able to meet deliveries in a matter of days, all stakeholders in the network must be aware of production times and phases, delivery dates, and components stock*". However, as was highlighted several times during the literature study section, companies frequently dislike working cooperatively with others since they are more concerned about losing market advantage by disclosing essential internal information. This subject repeatedly emerged during interviews, with experts stating that businesses outside the automotive industry frequently oppose these collaborations. When asked why businesses are always so wary of working together for the mutual benefit of both parties, participant "E" said that "*if businesses today are not even willing to share a maritime cargo container due to concerns about cargo contamination, sharing information that may or may not compromise certain business processes is a clear no-go*". According to some interviewees, some of these issues may be addressed by developing detailed plans for information-sharing mechanisms that would allow all parties to understand which information should remain private and which information should be shared with other participants, giving businesses the option to reveal important data, such as production timelines, stock levels, or delivery tracking, while concealing proprietary data, such as supplier contracts, sales figures, and other internal business-to-consumer metrics.

Companies should consider various security considerations when putting information sharing mechanisms in place within their supply chain, since privacy concerns are another important reason players don not cooperate or share information. As multiple scholars suggested, security measures like the implementation of Blockchain technology during data transactions would allow businesses to securely trade important information in a decertralised manner. When asked about the importance of Blockchain mechanisms in data transaction procedures, participants "B" and "C" acknowledged that "*businesses must realise the importance of all the procedures that help keep information secure, including encryption and the use of Blockchain, in order for them to feel secure using the information they have acquired and sharing it with other parties*". It's also interesting to consider the words of expert "J," as he pointed out that "*in Portugal, approximately 50% of companies still use paper to store certain critical information, rather than using cloud or physical servers*", offering an additional physical layer of security to various kinds of business information rather than implementing Blockchain measures.

Another widely discussed subject was using information systems in conjunction with Artificial Intelligence and Machine Learning to coordinate business activities better, allowing organisations to visualise graphs and product quantities on a single platform rather than depending on manual inventories or database consultations. Various experts made comparisons between the reality of past activities and the reality of present ones, explaining that a normal-sized storage facility can no longer function without a proper Warehouse Management System because businesses now demand a higher level of organisation and efficiency. In contrast, in the past, a single excel sheet would have been more than sufficient to record purchase orders, sales orders, and product quantity inside a small warehouse. Of course, the size of the organisation will always play a factor in this, since small enterprises with low income cannot afford to deploy a highly specialised information system and must instead rely on a less complex one or human labor, as mentioned by experts "I" and "J", where "*one thing is managing a single truck, and another is managing a fleet of vehicles*" and "*even if some businesses can run well without one, it is crucial to have an information system that allows us to manage expenses and requirements in day-to-day operations, especially if we are talking about large businesses*".

Companies may automate conventional business procedures by combining Machine Learning algorithms with Artificial Intelligence techniques. This may be particularly intriguing when businesses work with one or more partners. Supply chain visibility is still a major problem in Business to Business (B2B) transactions because stakeholders frequently can not trace items or materials while shipping between warehouses. Participant "O" considers that both Business to Consumer and Business to Business interactions require the same level of experience, claiming that "*it is illogical that, in B2C transactions, the consumer can easily track where the goods they have purchased are at any given time, whereas in B2B transactions, the company is unable to determine the location of a specific container holding tons of the same goods*". The same principle can be applied to a variety of industries, with many experts emphasising the importance of automation in information systems used in the food industry due to stringent regulations, as companies must always be aware of specifics regarding each product, including its origin (country, area, supplier, farm, and even individual animal or tree), packaging date, and expiration date.

Supply chain automation is essential for organisations to function at scale. For instance, businesses can create more items faster and with a higher level of consistency by automating industrial production operations. Due to the "Amazon effect," which has caused customers to anticipate quick delivery, retailers at the front end of the supply chain frequently face more significant challenges on last-mile deliveries. For these other stores to compete, automation helps to accelerate the delivery process, with many experts recognising that this kind of automation "*will aid companies to generate revenue in a world where competition will be fierce*". Automation may effectively shorten lead times by reducing waste and waiting periods. Automation of the process efficiently cuts down on machine downtime. It ensures raw materials are always available when needed, with expert "M" ensuring that "*the quicker material flow can guarantee quick customer deliveries, quick responses to demand changes, low capital ties, and quick detection of quality issues and other issues*".

"*Automation is frequently viewed as a way to save time and money, and this is undoubtedly true*", according to expert "P". By automating manual tasks, machines can perform activities that they are proficient at and nearly always more promptly than even the most highly skilled human. As a result, firms may spend less on workers to do routine administrative and data entry activities, freeing competent employees to concentrate on jobs that demand more brains than a computer can manage. Consistency, accuracy, and reproducibility are three of automation's main advantages. The system will continuously carry out a workflow after it has been automated. Automation is flexible if the correct procedures can be implemented properly before allowing the platform to take over, with expert "H" reminding businesses "*to hire qualified workers to properly codify automated algorithms and enable computers to automate labor-intensive tasks adequately*".

In summary, it is reasonable to claim that most of the interviews verified most of the concepts discovered through the literature study, with slight opinion differences based on the professionals' areas of expertise. One of the most significant considerations to verify relates to the original research objective, which was to determine whether (and how) Big Data Analytics may affect contemporary supply chains, easing the company's pain points and offering organisations improved methods for decision-making.

3.5.1 Supply chain visibility

According to academics and industry professionals, the absence of visibility throughout all business operations is one of the major problems with current supply chains. Businesses may encounter shortages, delays, or bottlenecks that eventually lead to higher expenses and revenue losses if they cannot track their inventory or check the status of their orders. Organisations face the danger of being unable to respond to shifting client demand, unforeseen interruptions or delays, and ultimately not delivering on their commitments without a comprehensive picture of the whole supply chain. Big Data Analytics may assist with resolving these problems by providing visibility into micro-level factors that directly affect product quality, stock management, warehouse efficiency, order fulfilment, customer needs, and market trends.

Product quality

Analytical tools can provide real-time insights on anticipated production effort and raw material requirements, labor and machine deployment timetables, and coordinated logistics to assist a manufacturing plant in running as efficiently as possible. The most viable candidates for optimisation, whether processes, machinery, or personnel schedules, may be found by analysing data patterns inside processes and workflows. This will ensure that production output is raised while quality assurance is maintained.

Inventory management

Predictive data may help businesses identify market patterns using Big Data Analytics. This information, which identifies shopping trends or behaviours, is frequently taken from inventory management histories. Data may also show the motivations behind each purchase and why particular goods are returned. As a result, market seasonality and patterns may be monitored and used to forecast future trends that would be similar. Managers will thus know what to fill the shelves with or prepare for the next holiday season. Since the data is frequently changing, businesses may also learn what precise pricing should be utilised for each market group for each product. A business would benefit greatly from being one step ahead of the competition since it could meet client demand as soon as it arose.

Warehouse efficiency

A further factor to take into account is efficient logistics. The delivery procedure is even more crucial for internet businesses, whose main activity is selling remotely. Orders are often instantly assigned to specific warehouses to save shipment times and other logistical expenses. Based on past purchases, businesses can forecast which products will be in high demand across various markets and ensure they are always in stock. As a result, planning techniques can increase consumer happiness through a quicker shopping experience, cheaper expenses, and product availability

3.5.2 Data standardisation

Data uniformization is a major problem for modern supply chains, even though Big Data cannot necessarily solve it. Businesses that rely on multiple information systems struggle to get a straightforward output from each system because they all use different languages, inputs, and information processing methods. For multi-tier supply chains, standardisation must be implemented for businesses to accurately measure what is happening throughout the chain. In the long term, standardisation is also cost-efficient, as businesses will not have to start over every time after data treatment standards have been established. One may take the first step toward establishing a standardised supply chain by examining current procedures and resources. Increased efficiency should result from streamlining procedures, which, once established, will work silently in the background to improve the overall efficacy of any given activity.

3.5.3 Automation

Regardless of the industry, automation is a vital component of contemporary enterprises. While the retail and automotive sectors are already highly automated, with a variety of Artificial Intelligence and Machine Learning algorithms in place to assist businesses in reducing their employees' manual work (while also reducing the associated human risk), sectors like transport logistics and maritime deliveries still have a long way to go due to a significant lack of digital and automated developments.

In this case, some specialists on harbour operations concurred that Big Data would provide ports more direct access to statistical data, enabling the transformation of some processes that are currently highly rudimentary (in comparison to other more modern seaports).

Maritime activities

Big Data may be used to control ship sensors and do predictive analysis, necessary to stop delays and boost the industry's overall operational effectiveness. Accurate cargo tracking is crucial for the shipping sector to maintain the requisite safety and secrecy. Information on the causes of vessel losses at sea, losses of containers within or outside terminals or warehouses, and other dispatch-related issues (for example, the causes of goods damage) may be gathered through data obtained through appropriate tracking of shipments over several years. Big Data for the shipping sector may be utilised to guide future decisions to anticipate and avert pricey issues and provide more dependable solutions for cargo transportation. Experts concur that Big Data Analytics might be crucial to ship design. The suggested ship design might be tested using these older datasets without physically building it. This is a massive benefit for the shipbuilding industry. The need for cargo transportation and related logistical assistance will increase exponentially as the world transitions to a more globally integrated economy. To have the most lucrative delivery methods, there will eventually be a more significant requirement to maximise time and profit, meaning that the delivery of products will be more effective thanks to the employment of cutting-edge data processing tools.

Future of automation

Big Data is assisting businesses in fully automating all daily processes. A new type of intelligent manufacturing technology is now possible due to the development of networking technologies and the growth of the Internet of Things. Big Data Analytics may boost accuracy and yield, and speed up assembly in the manufacturing business, which will assist in decreasing waste and energy expenses. Big Data will be shaped for future trends using augmented analytics. These technologies may improve analysts' work speed, effectiveness, and accuracy. By lowering technological barriers to analysis and making more complex approaches accessible to those with less developed data skills and expertise, Machine Learning and natural language technologies also aid in bringing domain experts closer to their data. Business users and executives could benefit from augmented analytics because these technologies would enable them to swiftly extract value from their data without requiring in-depth technical knowledge or data experience.

Businesses will also gain various advantages from the development and interoperability of Big Data and the Internet of Things. Big Data Analytics assists in making sense of the data and information that Internet of Things devices collect. The massive amounts of unstructured data that have been gathered are organised into smaller data sets using these technologies, which may help businesses get insights into how their operations perform and enhance decision-making. When combined with the Internet of Things, descriptive, diagnostic, and prescriptive analytics are just a few insights that Big Data Analytics may offer.

3.5.4 Digital transformation

The interviewed professionals also confirmed the significance of digitisation in today's industries. Companies cannot deploy automated solutions if their operations are not entirely digitised since most of their information is still on paper or in other manual tasks. Companies undergoing digital transformation can better adapt to change and maintain their competitiveness in an increasingly digital environment. The opportunity for a business to integrate the two in their efforts is where Big Data plays a role in digital transformation, allowing the digitisation and automation of corporate processes. This automation and digitisation could help companies increase productivity, promote creativity, and create new business models. Additionally, Big Data Analytics enables firms to view detailed data about certain or several client groups. This can provide information about their online activities, the things they purchase, how frequently they do so, and if they plan to continue purchasing those same products. Businesses may thus use this knowledge to make changes to better serve their consumers' future needs while also establishing goals for how to do so.

3.6 Vital Supply Chain indicators

Supply chain performance indicators are essential for monitoring the ongoing growth, success, evolution and development of any company's supply, fulfilment, and delivery activities. Business managers may identify inefficiencies within the company's ecosystem by gathering, curating, and evaluating critical supply chain indicators, while also being able to capitalise on the ecosystem's present strengths and set goals that will assist their supply chain grow along with the company's success. By reviewing the key findings from the literature review and what was learned from the interviews, it is possible to develop a set of supply chain indicators that could aid business managers in keeping track of specific metrics that might instinctively indicate performance growth or decay in routine day-to-day business operations. The development of a virtual dashboard that could integrate well-known Key Performance Indicators, such as perfect order rate, inventory turnover, warehousing costs, and delivery time, with new global performance metrics, such as degree of stakeholder collaboration, level of supply chain visibility, and degree of automation in manufacturing, warehousing, and shipping processes, could significantly boost business performance.

3.6.1 Supply Chain collaboration

Long-term connections eventually result in increased revenues, and the benefits of supply chain collaboration are similar to the advantages of customer retention. This occurs, among other things, because business partners can better synchronise quality production and ethical working standards throughout their supply chains when enough time has been spent working together and knowing one another. Additionally, having stronger working relationships with suppliers and customers enables businesses to maximise productivity while minimising the wastage of resources like time and money. Consequently, it would be ideal to implement mechanisms to gauge stakeholder cooperation across the supply chain by exchanging vital information and preserving a degree of engagement that could be regularly evaluated.

Businesses could have an early indication of how effective a partnership is by analysing simple, upstream metrics like the number of cooperative projects launched, the overall success rates of projects implemented, and the number of patent applications submitted over time. As projects progress, data will assist in increasing the accuracy and usefulness of KPIs like the expected return on investment.

Collaboration projects are meant to provide value to the top line, but collaborative connections may greatly impact the performance of the bottom line. This is valid for activities that increase production, promote efficiency, and cut waste, such as the redesign of production facilities, the preparation of new supply schedules and transportation routes, and the adoption of ERPs that might give partners better visibility into the quantity of stock, deliveries, and needs at a certain moment. In these instances, creating a single dashboard that includes several data for key procurement tasks, such as the amount of money saved after adopting particular initiatives, supplier lead time, rate of emergency purchases, and vendor availability might be a helpful tool for businesses to evaluate the indirect effects of working with a specific supply chain partner.

Measuring the productivity of collaboration initiative portfolios is a more sophisticated method of evaluating the effectiveness of supplier collaboration since upstream measures like those stated above are insightful and valuable as early indicators, but they don't really show how well supplier collaboration projects are being monetised. Creating indicators for collaborative projects that provide an easy way to monitor conversion (such as concept to product) offers a solid sense of the program's overall performance. Key metrics, such as the ratio of gross margin to new product sales or the ratio of research and development expenditure to the number of new product sales over time, which show whether the products developed through supplier collaboration are profitable or not, could help businesses understand the results of partner collaborations and help managers determine whether the partnership is beneficial or not.

3.6.2 Supply Chain visibility

Just as attaining a high degree of visibility can be a significant operational challenge, measuring visibility can be somewhat of a hurdle in its own right. Effective Supply Chain Management depends on the supply chain's visibility level. The key to avoiding interruptions, managing costs, providing products and services to consumers, and making a profit for the firm is understanding the current condition of every element inside the process. Global supply networks have been disrupted since the outbreak began in 2020, and these disruptions are still going strong. According to several interviews, this wake-up call is enticing business managers to concentrate more on enhancing end-to-end supply chain visibility to attain greater resilience, improve performance, and reach higher supply chain maturity stages. However, organisations are discovering that visibility gaps within production facilities, distribution hubs, and across locations are hampering their strategic or digital transformation. Many business and operational procedures that connect locations in the ecosystem today miss opportunities to record crucial multidimensional data, such as location-specific edge data, temperature, humidity, shock, and other condition data essential for advanced planning and execution. As a result, companies should adopt certain best-practice strategies that could aid in assessing the degree of visibility of their supply chains, pinpointing potential pain areas, and obtaining a clear image of where the supply chain stands.

The on-time delivery rate is one of the most crucial metrics for supply chain visibility. This indicator counts the proportion of orders or shipments that arrive within the predetermined or anticipated time range. A high percentage of deliveries made on time shows that the supply chain is effective, dependable, and attentive to client demands. Low on-time delivery rates indicate that the supply chain is experiencing hiccups, delays, or mistakes that negatively impact the customer experience and satisfaction. Monitoring and optimising the supply chain's partners, resources, and processes are required to raise the percentage of deliveries that arrive on time.

The visibility of shipment status is another important metric for supply chain visibility. This indicator gauges the degree to which the supply chain can access and distribute real-time details regarding the whereabouts, state, and status of shipments while they are in route. Strong visibility of the shipping status shows that the supply chain can track and monitor the performance of shipments and react promptly and effectively to any problems or changes. Companies may be exposed to uncertainties, challenges, or inefficiencies in the delivery process if there is little or no insight into the shipments' status. Innovative technologies like Global Positioning System (GPS), Radio frequency identification (RFID), Internet of Things (IoT), or Blockchain could be used by businesses better to understand the status of their cargo throughout transportation processes.

Controlling and monitoring visibility (and security) across the supply network may also be accomplished by combining the previous indicator with an index that tracks occurrences involving shipment security. This indicator measures the occurrence and seriousness of events that compromise the security of shipments in transit. These instances might involve theft, shipment damage, loss, diversion, tampering, or sabotage. A low shipment security incident index indicates the supply chain's high degree of security and resilience and ability to shield shipments from internal and external threats. A supply chain with low security and resilience and a high number of shipment security incidents may have financial, reputational, or legal repercussions.

Closing visibility gaps is now achievable thanks to technological advancements that allow for collecting precise data at every stage of the supply chain. This thorough data collection enables a far broader, more comprehensive data flow to the cloud and crucial planning and monitoring systems. The planning systems, in which a significant amount of investment has been made to date, may now depend on data that spans the whole supply chain rather than a small number of points or just one section of the supply chain. The potential for significantly better accuracy in planning and monitoring systems is delivered by increased visibility, making it possible to apply predictive analytics to foresee and prevent problems.

3.6.3 Supply Chain automation

These days, Supply Chain Management isn't just about production and delivery; it's also about improving the entire customer experience. However, some obstacles prevent organisations from evolving and cause gaps and vulnerabilities in supply chains, diverting the company's attention from raising the level of its service quality to addressing other issues. Following manual procedures and delaying automation plans cause many of these difficulties. As was previously said, a critical challenge in implementing Big Data Analytics and automating the supply chain is the large quantity of data that businesses gather, handle, and evaluate to obtain relevant information. However, all of this data must be handled within a certain time period; otherwise, it will go out-of-date and cost the organisation money and effort while allowing it to draw almost no conclusions or outcomes. Supply chain automation enables businesses to carry out these operations using technology or algorithms, freeing up skilled individuals to perform other jobs that demand human labor and are less prone to mistakes.

Companies may monitor certain Key Performance Indicators that help them gauge the degree of success of their digital transformation by implementing Machine Learning and Artificial Intelligence technologies to automate the majority of their manual labour, like the time saved compared to other manual tasks, the number of errors made during an automated workflow or process, or the amount of money the automated processes have saved. These metrics could help businesses understand their level of digitisation, along with less obvious ones like the proportion of paper-less operations, the degree of digital manufacturing operations, and the amount of information still stored outside of virtual information softwares like ERP and WMS systems, requiring executives to take on new projects to change their supply chain into one that is more digital and flexible, which not only lowers costs but also adds value for the organisation and its employees.

Figure 3.3 provides a visual breakdown of all the discussed supply chain metrics, organised by category.

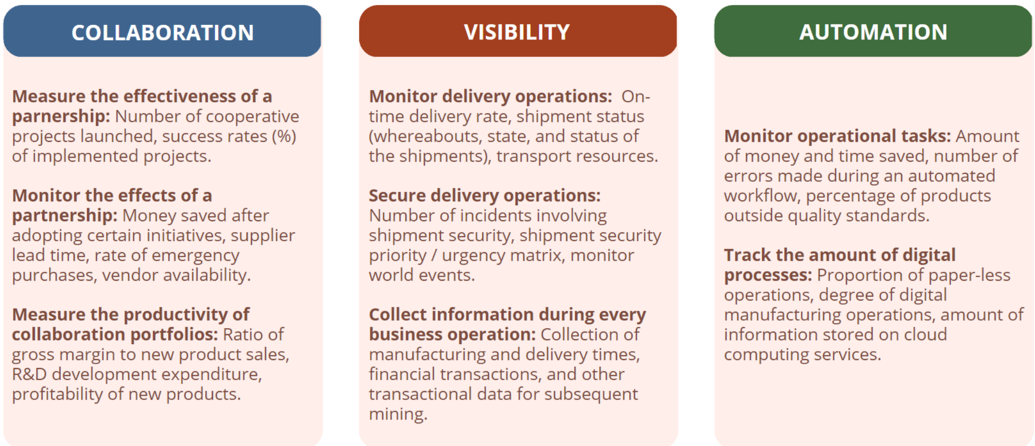


Figure 3.3: Vital Supply Chain indicators.

Chapter 4

Conclusion

The research scope of this dissertation focuses on how Big Data Analytics could revolutionise how firms engage with the global supply chain. Namely, given its significant effect on business processes, supply chain agility and visibility, and, most crucially, customer relationships. Big Data is becoming increasingly important in Supply Chain Management. With its real-time analytics capabilities, Big Data can address various strategic, operational, and tactical issues, tremendously influencing all supply chain processes, from procurement to delivery. Product traceability is essential for efficient supply chain operations. Big Data Analytics allows companies to collect precise product information, allowing operators to keep up with their distribution cycle. The improved traceability guarantees that items are tracked from manufacture to sale. Improved traceability will enable organisations to cooperate better with supply chain partners to simplify distribution.

Artificial Intelligence and Machine Learning techniques also deliver powerful optimisation skills necessary for more accurate capacity planning, increased productivity, high quality, reduced costs, and higher product output. However, like many other technologies, AI and ML have limitations. Human-driven decision-making will occasionally surpass model-based methods. In addition, if humans do not interfere, software systems may suggest solutions that meet the constraints but at an unacceptable cost or risk, forcing companies to rethink their strategies. Data-driven AI supply chain management solutions outperform traditional Artificial Intelligence and human-driven systems. However, supply chain managers should retain their power and sense of responsibility for carrying out the recommendations of these systems. Similarly, software engineers should be more compelling in teaching supply chain managers about the nature of the data they are using, the limitations of the systems, and how to assess the accuracy and usefulness of the systems' solutions over time.

It is foreseeable that the data environment will continue to expand in complexity as new digital technologies generate new data streams to supplement the current plethora. Artificial Intelligence, virtual reality, and speech recognition are just three areas of innovation that are rapidly growing. The issue of preserving data quality is already massive. Researchers claim that poor data quality causes millions of dollars in yearly losses. As a result, effective data cleansing will become increasingly crucial. It's only time before all personal datasets become old and erroneous. However, the rate at which the firms' data decays will be determined by various factors, ranging from the consumers or sectors they service to the regions where they do business.

Multiple interviews were conducted on several of these theoretical notions and ideas, and experts validated most of the conclusions from the literature research. The interviews discovered that many of the given hypotheses made sense in modern supply chains, despite many selected publications lacking even a single real-life case study. As Big Data is giving supplier networks better data accuracy, clarity, and insights, leading to more contextual knowledge shared across every organisation, experts agreed that Big Data would alter how businesses engage with partners throughout their supply chain. Several examples of Big Data Analytics applications in sales, inventory, and operations planning were discussed during these interviews, with participants providing some insight into how their organisations approached data collection, treatment, and analysis procedures and the impact these procedures have on their business processes, whether they are financial or logistical.

Naturally, there is still a significant difference between every industry's development and automation. It became apparent when speaking with professionals from various automotive and retail firms that their reality differed from the one that academics or directors of marine companies had described. While businesses operating in the automotive and retail supply chains already had Big Data Analytics programs that were well-implemented and were gathering and processing lots of data to aid managers in making better choices and better controlling the company's KPI and overall expenses, firms engaged in marine distribution or harbour operations still lack some growth in this area. Even in 2023, some basic processes that could already be automated (and are already automated in other industries) were still carried out manually by employees, some even on paper and without using cloud computing services, were confirmed by multiple experts. This presents a challenge when it comes to data sharing and data analysis.

The number of publications chosen for evaluation and the number of individuals contacted certainly limit the validation and discovery of novel concepts about Big Data Analytics applications. New publications might be able to offer some fresh perspectives on how these strategies are now being used, even if the chosen papers were among the most referenced and authored by well-known scholars who frequently share their work on the subject. The primary industry of each participant also impacts the findings of this study. The fact that many interviewees worked in fields with advanced technological development meant that the majority of them were already utilising Big Data applications daily, whether they were fully packaged solutions with information collection, treatment, analysis, and sharing stages or by simply gathering data from external public sources and using it in simple ML algorithms to retain some knowledge and predict future trends. There was still room for improvement in this particular area, as some participants held senior positions within their organisations and worked on the more practical side of the business, lacking some technological concepts that would help provide programmers more insight into how to implement Big Data Analytics solutions better when combined with Artificial Intelligence and Machine Learning. A future extension of this research may include more technical insights and information about the effects of the physical internet on contemporary supply chains and how Blockchain may significantly alter how companies communicate with one another and with customers.

There is still a substantial gap in the literature review about the application and usefulness of Big Data in various organisations. While Big Data Analytics appears to be the ideal technique for propelling enterprises forward, there is a shortage of study cases that truly examine the results of applying Big Data to firms from the ground up. Thus, future studies should analyse Big Data Analytics (BDA)' capacity to increase firm efficiency and effectiveness (e.g., bottleneck detection, enhanced predictive maintenance, and scenario development for improved quality control). Little research also links Augmented Data Management and Physical Internet techniques to Big Data solutions. Physical Internet solutions would provide businesses access to a network of routes connected by hubs and various means of transportation, allowing them to expedite how items are moved from one location to another. Since many organisations operate essentially autonomously, with no shared warehouses or transportation routes, the notion of opening them up to everyone is still debated by many, as shown in multiple interviews, since corporations seek a reasonable ground between sharing information and maintaining complete control of their market position.

References

[1] C. R. Carter and P. L. Easton, "Sustainable supply chain management: evolution and future directions," *International Journal of Physical Distribution & Logistics Management*, vol. 41, pp. 46–62, feb 2011. [Cited on page 1]

[2] S. Mohr and O. Khan, "3d printing and its disruptive impacts on supply chains of the future," *Technology Innovation Management Review*, vol. 5, pp. 20–25, nov 2015. [Cited on page 1]

[3] H. Min and G. Zhou, "Supply chain modeling: past, present and future," *Computers & Industrial Engineering*, vol. 43, pp. 231–249, jul 2002. [Cited on page 1]

[4] S. Min, Z. G. Zacharia, and C. D. Smith, "Defining supply chain management: In the past, present, and future," *Journal of Business Logistics*, vol. 40, pp. 44–55, mar 2019. [Cited on page 1]

[5] S. J. Hu, "Evolving paradigms of manufacturing: From mass production to mass customization and personalization," *Procedia CIRP*, vol. 7, pp. 3–8, 2013. [Cited on page 1]

[6] G. Richards, *Warehouse management a complete guide to improving efficiency and minimizing costs in the modern warehouse.* 2018. [Cited on page 1]

[7] D. Guerrero and J.-P. Rodrigue, "The waves of containerization: shifts in global maritime transportation," *Journal of Transport Geography*, vol. 34, pp. 151–164, jan 2014. [Cited on page 1]

[8] T. Wang and K. Cullinane, "The efficiency of european container terminals and implications for supply chain management," in *Port Management*, pp. 253–272, Palgrave Macmillan UK, 2015. [Cited on page 2]

[9] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?," *Technological Forecasting and Social Change*, vol. 114, pp. 254–280, jan 2017. [Cited on page 2]

[10] Z. Seyedghorban, H. Tahernejad, R. Meriton, and G. Graham, "Supply chain digitalization: past, present and future," *Production Planning & Control*, vol. 31, pp. 96–114, dec 2019. [Cited on page 2]

[11] Y. fen Su and C. Yang, "Why are enterprise resource planning systems indispensable to supply chain management?," *European Journal of Operational Research*, vol. 203, pp. 81–94, may 2010. [Cited on page 2]

[12] H. Ince, S. Z. Imamoglu, H. Keskin, A. Akgun, and M. N. Efe, "The impact of ERP systems and supply chain management practices on firm performance: Case of turkish companies," *Procedia - Social and Behavioral Sciences*, vol. 99, pp. 1124–1133, nov 2013. [Cited on page 2]

[13] G. Wang, A. Gunasekaran, E. W. Ngai, and T. Papadopoulos, "Big data analytics in logistics and supply chain management: Certain investigations for research and applications," *International Journal of Production Economics*, vol. 176, pp. 98–110, jun 2016. [Cited on pages 2 and 25]

[14] D. Ivanov, A. Dolgui, and B. Sokolov, "The impact of digital technology and industry 4.0 on the ripple effect and supply chain risk analytics," *International Journal of Production Research*, vol. 57, pp. 829–846, jun 2018. [Cited on pages 2 and 30]

[15] J. Wolf, "The relationship between sustainable supply chain management, stakeholder pressure and corporate sustainability performance," *Journal of Business Ethics*, vol. 119, pp. 317–328, feb 2013. [Cited on page 2]

[16] A. P. Barbosa-Póvoa, C. da Silva, and A. Carvalho, "Opportunities and challenges in sustainable supply chain: An operations research perspective," *European Journal of Operational Research*, vol. 268, pp. 399–431, jul 2018. [Cited on page 2]

[17] S. R. Cardoso, A. P. F. Barbosa-Póvoa, and S. Relvas, "Design and planning of supply chains with integration of reverse logistics activities under demand uncertainty," *European Journal of Operational Research*, vol. 226, pp. 436–451, may 2013. [Cited on page 2]

[18] Chen, Chiang, and Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quarterly*, vol. 36, no. 4, p. 1165, 2012. [Cited on page 2]

[19] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *International Journal of Information Management*, vol. 35, pp. 137–144, apr 2015. [Cited on pages 2 and 24]

[20] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, and S. U. Khan, "The rise of "big data" on cloud computing: Review and open research issues," *Information Systems*, vol. 47, pp. 98–115, jan 2015. [Cited on page 2]

[21] P. N. Wognum, H. Bremmers, J. H. Trienekens, J. G. van der Vorst, and J. M. Bloemhof, "Systems for sustainability and transparency of food supply chains – current status and challenges," *Advanced Engineering Informatics*, vol. 25, pp. 65–76, jan 2011. [Cited on page 2]

[22] T. Gardner, M. Benzie, J. Börner, E. Dawkins, S. Fick, R. Garrett, J. Godar, A. Grimard, S. Lake, R. Larsen, N. Mardas, C. McDermott, P. Meyfroidt, M. Osbeck, M. Persson, T. Sembres, C. Suavet, B. Strassburg, A. Trevisan, C. West, and P. Wolvekamp, "Transparency and sustainability in global commodity supply chains," *World Development*, vol. 121, pp. 163–177, sep 2019. [Cited on page 2]

[23] M. J. Meixell and P. Luoma, "Stakeholder pressure in sustainable supply chain management," *International Journal of Physical Distribution & Logistics Management*, vol. 45, pp. 69–89, mar 2015. [Cited on page 2]

[24] Q. Zhu, J. Sarkis, and K. hung Lai, "Examining the effects of green supply chain management practices and their mediations on performance improvements," *International Journal of Production Research*, vol. 50, pp. 1377–1394, mar 2012. [Cited on page 3]

[25] M. S. Sodhi and C. S. Tang, "Research opportunities in supply chain transparency," *Production and Operations Management*, vol. 28, pp. 2946–2959, oct 2019. [Cited on page 3]

[26] H. Min and I. Kim, "Green supply chain research: past, present, and future," *Logistics Research*, vol. 4, pp. 39–47, feb 2012. [Cited on page 3]

[27] M.-L. Tseng, M. S. Islam, N. Karia, F. A. Fauzi, and S. Afrin, "A literature review on green supply chain management: Trends and future challenges," *Resources, Conservation and Recycling*, vol. 141, pp. 145–162, feb 2019. [Cited on page 3]

[28] P. Dutta, T.-M. Choi, S. Somani, and R. Butala, "Blockchain technology in supply chain operations: Applications, challenges and research opportunities," *Transportation Research Part E: Logistics and Transportation Review*, vol. 142, p. 102067, oct 2020. [Cited on page 3]

[29] F. Kache and S. Seuring, "Challenges and opportunities of digital information at the intersection of big data analytics and supply chain management," *International Journal of Operations & Production Management*, vol. 37, pp. 10–36, jan 2017. [Cited on page 3]

[30] S. Zhu, J. Song, B. T. Hazen, K. Lee, and C. Cegielski, "How supply chain analytics enables operational supply chain transparency," *International Journal of Physical Distribution & Logistics Management*, vol. 48, pp. 47–68, jan 2018. [Cited on page 3]

[31] R. Srinivasan and M. Swink, "An investigation of visibility and flexibility as complements to supply chain analytics: An organizational information processing theory perspective," *Production and Operations Management*, vol. 27, pp. 1849–1867, aug 2017. [Cited on page 3]

[32] D. Mishra, A. Gunasekaran, T. Papadopoulos, and R. Dubey, "Supply chain performance measures and metrics: a bibliometric study," *Benchmarking: An International Journal*, vol. 25, pp. 932–967, apr 2018. [Cited on page 3]

[33] D. Prajogo and J. Olhager, "Supply chain integration and performance: The effects of long-term relationships, information technology and sharing, and logistics integration," *International Journal of Production Economics*, vol. 135, pp. 514–522, jan 2012. [Cited on page 3]

[34] S. E. Fawcett, C. Wallin, C. Allred, A. M. Fawcett, and G. M. Magnan, "Information technology as an enabler of supply chain collaboration: A dynamic-capabilities perspective," *Journal of Supply Chain Management*, vol. 47, pp. 38–59, jan 2011. [Cited on page 3]

[35] D. Arunachalam, N. Kumar, and J. P. Kawalek, "Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice," *Transportation Research Part E: Logistics and Transportation Review*, vol. 114, pp. 416–436, jun 2018. [Cited on page 3]

[36] R. Y. Zhong, S. T. Newman, G. Q. Huang, and S. Lan, "Big data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives," *Computers & Industrial Engineering*, vol. 101, pp. 572–591, nov 2016. [Cited on page 3]

[37] M. Saunders, P. Lewis, A. Thornhill, and A. Bristow, *"Research Methods for Business Students" Chapter 4: Understanding research philosophy and approaches to theory development*, pp. 128–171. 03 2019. [Cited on page 4]

[38] M. Templier and G. Paré, "A framework for guiding and evaluating literature reviews," *Communications of the Association for Information Systems*, vol. 37, 2015. [Cited on page 7]

[39] M. A. Waller and S. E. Fawcett, "Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management,"

Journal of Business Logistics, vol. 34, pp. 77–84, jun 2013. [Cited on pages 24 and 26]

[40] A. Rajeev, R. K. Pati, S. S. Padhi, and K. Govindan, "Evolution of sustainability in supply chain management: A literature review," *Journal of Cleaner Production*, vol. 162, pp. 299–314, sep 2017. [Cited on page 24]

[41] D. M. Lambert, *Fundamentals of logistics management.* Irwin/McGraw-Hill, 1998. [Cited on page 24]

[42] Y. Tribis, A. E. Bouchti, and H. Bouayad, "Supply chain management based on blockchain: A systematic mapping study," *MATEC Web of Conferences*, vol. 200, p. 00020, 2018. [Cited on page 24]

[43] T. Rupasinghe, "Internet of things (iot) embedded future supply chains for industry 4.0: An assessment from an erp-based fashion apparel and footwear industry," *Journal of Supply Chain Management*, vol. 6, 01 2017. [Cited on page 24]

[44] G. Gereffi and J. Lee, "Why the world suddenly cares about global supply chains," *Journal of Supply Chain Management*, vol. 48, pp. 24–32, jul 2012. [Cited on page 24]

[45] A. G. Frank, L. S. Dalenogare, and N. F. Ayala, "Industry 4.0 technologies: Implementation patterns in manufacturing companies," *International Journal of Production Economics*, vol. 210, pp. 15–26, apr 2019. [Cited on page 24]

[46] A. Gunasekaran, T. Papadopoulos, R. Dubey, S. F. Wamba, S. J. Childe, B. Hazen, and S. Akter, "Big data and predictive analytics for supply chain and organizational performance," *Journal of Business Research*, vol. 70, pp. 308–317, jan 2017. [Cited on page 24]

[47] R. Geng, S. A. Mansouri, and E. Aktas, "The relationship between green supply chain management and performance: A meta-analysis of empirical evidences in asian emerging economies," *International Journal of Production Economics*, vol. 183, pp. 245–258, jan 2017. [Cited on page 24]

[48] R. Dubey, A. Gunasekaran, D. J. Bryde, Y. K. Dwivedi, and T. Papadopoulos, "Blockchain technology for enhancing swift-trust, collaboration and resilience within a humanitarian supply chain setting," *International Journal of Production Research*, vol. 58, pp. 3381–3398, feb 2020. [Cited on page 24]

[49] J. Liebowitz, ed., *Big Data and Business Analytics.* Auerbach Publications, apr 2016. [Cited on page 24]

[50] R. Vidgen, S. Shaw, and D. B. Grant, "Management challenges in creating value from business analytics," *European Journal of Operational Research*, vol. 261, pp. 626–639, sep 2017. [Cited on page 24]

[51] S. Tiwari, H. Wee, and Y. Daryanto, "Big data analytics in supply chain management between 2010 and 2016: Insights to industries," *Computers & Industrial Engineering*, vol. 115, pp. 319–330, jan 2018. [Cited on page 24]

[52] D. Ivanov, "Viable supply chain model: integrating agility, resilience and sustainability perspectives lessons from and thinking beyond the covid-19 pandemic," *Annals of Operations Research*, vol. 319, pp. 1411–1431, may 2020. [Cited on page 24]

[53] L. Duan and Y. Xiong, "Big data analytics and business analytics," *Journal of Management Analytics*, vol. 2, pp. 1–21, jan 2015. [Cited on page 24]

[54] G. Phillips-Wren, L. S. Iyer, U. Kulkarni, and T. Ariyachandra, "Business analytics in the context of big data: A roadmap for research," *Communications of the Association for Information Systems*, vol. 37, 2015. [Cited on page 25]

[55] S. Mithas, M. R. Lee, S. Earley, S. Murugesan, and R. Djavanshir, "Leveraging big data and business analytics [guest editors' introduction]," *IT Professional*, vol. 15, pp. 18–20, nov 2013. [Cited on page 25]

[56] B. K. Chae, "Insights from hashtag # supplychain and twitter analytics: Considering twitter and twitter data for supply chain practice and research," *International Journal of Production Economics*, vol. 165, pp. 247–259, jul 2015. [Cited on page 25]

[57] S. Ren, Y. Zhang, Y. Liu, T. Sakao, D. Huisingh, and C. M. Almeida, "A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions," *Journal of Cleaner Production*, vol. 210, pp. 1343–1365, feb 2019. [Cited on page 25]

[58] D. Q. Chen, D. S. Preston, and M. Swink, "How the use of big data analytics affects value creation in supply chain management," *Journal of Management Information Systems*, vol. 32, pp. 4–39, oct 2015. [Cited on page 25]

[59] R. Dubey, A. Gunasekaran, S. J. Childe, C. Blome, and T. Papadopoulos, "Big data and predictive analytics and manufacturing performance: Integrating institutional theory, resource-based view and big data culture," *British Journal of Management*, vol. 30, pp. 341–361, apr 2019. [Cited on page 25]

[60] Y. Liu, "Big data and predictive business analytics," *Journal of Business Forecasting*, vol. 33, p. 40, 2014. [Cited on page 25]

[61] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," *IEEE Access*, vol. 2, pp. 652–687, 2014. [Cited on page 25]

[62] T. Schoenherr and C. Speier-Pero, "Data science, predictive analytics, and big data in supply chain management: Current state and future potential," *Journal of Business Logistics*, vol. 36, pp. 120–132, feb 2015. [Cited on page 25]

[63] B. T. Hazen, C. A. Boone, J. D. Ezell, and L. A. Jones-Farmer, "Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications," *International Journal of Production Economics*, vol. 154, pp. 72–80, aug 2014. [Cited on page 25]

[64] O. Kwon, N. Lee, and B. Shin, "Data quality management, data usage experience and acquisition intention of big data analytics," *International Journal of Information Management*, vol. 34, pp. 387–394, jun 2014. [Cited on page 25]

[65] L. D. Xu, "Information architecture for supply chain quality management," *International Journal of Production Research*, vol. 49, pp. 183–198, jan 2011. [Cited on page 25]

[66] A. Katal, M. Wazid, and R. H. Goudar, "Big data: Issues, challenges, tools and good practices," *2013 Sixth International Conference on Contemporary Computing (IC3)*, pp. 404–409, 2013. [Cited on page 25]

[67] S. Sagiroglu and D. Sinanc, "Big data: A review," in *2013 International Conference on Collaboration Technologies and Systems (CTS)*, IEEE, may 2013. [Cited on page 25]

[68] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big data: Issues and challenges moving forward," in *2013 46th Hawaii International Conference on System Sciences*, IEEE, jan 2013. [Cited on page 25]

[69] S. F. Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou, "How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study," *International Journal of Production Economics*, vol. 165, pp. 234–246, jul 2015. [Cited on page 25]

[70] R. Y. Zhong, C. Xu, C. Chen, and G. Q. Huang, "Big data analytics for physical internet-based intelligent manufacturing shop floors," *International Journal of Production Research*, vol. 55, pp. 2610–2621, sep 2015. [Cited on page 25]

[71] R. Y. Zhong, G. Q. Huang, S. Lan, Q. Dai, X. Chen, and T. Zhang, "A big data approach for logistics trajectory discovery from RFID-enabled production data," *International Journal of Production Economics*, vol. 165, pp. 260–272, jul 2015. [Cited on page 25]

[72] E. Hofmann, "Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect," *International Journal of Production Research*, vol. 55, pp. 5108–5126, jul 2015. [Cited on page 25]

[73] W. A. Günther, M. H. R. Mehrizi, M. Huysman, and F. Feldberg, "Debating big data: A literature review on realizing value from big data," *The Journal of Strategic Information Systems*, vol. 26, pp. 191–209, sep 2017. [Cited on page 25]

[74] D. Mishra, A. Gunasekaran, T. Papadopoulos, and S. J. Childe, "Big data and supply chain management: a review and bibliometric analysis," *Annals of Operations Research*, vol. 270, pp. 313–336, jun 2016. [Cited on page 25]

[75] L. Einav and J. Levin, "Economics in the age of big data," *Science*, vol. 346, nov 2014. [Cited on page 25]

[76] S. Rajput and S. P. Singh, "Connecting circular economy and industry 4.0," *International Journal of Information Management*, vol. 49, pp. 98–113, dec 2019. [Cited on page 25]

[77] S. F. Wamba, R. Dubey, A. Gunasekaran, and S. Akter, "The performance effects of big data analytics and supply chain ambidexterity: The moderating effect of environmental dynamism," *International Journal of Production Economics*, vol. 222, p. 107498, apr 2020. [Cited on page 26]

[78] A. Belhadi, S. Kamble, C. J. C. Jabbour, A. Gunasekaran, N. O. Ndubisi, and M. Venkatesh, "Manufacturing and service supply chain resilience to the covid-19 outbreak: Lessons learned from the automobile and airline industries," *Technological Forecasting and Social Change*, vol. 163, p. 120447, feb 2021. [Cited on page 26]

[79] G. Büyüközkan and F. Göçer, "Digital supply chain: Literature review and a proposed framework for future research," *Computers in Industry*, vol. 97, pp. 157–177, may 2018. [Cited on page 26]

[80] S. H. A. El-Sappagh, A. M. A. Hendawi, and A. H. E. Bastawissy, "A proposed model for data warehouse ETL processes," *Journal of King Saud University - Computer and Information Sciences*, vol. 23, pp. 91–104, jul 2011. [Cited on page 26]

[81] L. Mufioz, J. N. Mazon, and J. Trujillo, "ETL process modeling conceptual for data warehouses: A systematic mapping study," *IEEE Latin America Transactions*, vol. 9, pp. 358–363, jun 2011. [Cited on page 26]

[82] S. K. Bansal, "Towards a semantic extract-transform-load (ETL) framework for big data integration," in *2014 IEEE International Congress on Big Data*, IEEE, jun 2014. [Cited on page 26]

[83] K. Kakish and T. Kraft, "Etl evolution for real-time data warehousing," 2012. [Cited on page 26]

[84] D. E. O'Leary, "Artificial intelligence and big data," *IEEE Intelligent Systems*, vol. 28, pp. 96–99, mar 2013. [Cited on page 27]

[85] R. Dubey, A. Gunasekaran, S. J. Childe, D. J. Bryde, M. Giannakis, C. Foropon, D. Roubaud, and B. T. Hazen, "Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations," *International Journal of Production Economics*, vol. 226, p. 107599, aug 2020. [Cited on page 27]

[86] G. Baryannis, S. Validi, S. Dani, and G. Antoniou, "Supply chain risk management and artificial intelligence: state of the art and future research directions," *International Journal of Production Research*, vol. 57, pp. 2179–2202, oct 2018. [Cited on page 27]

[87] L. Zhou, S. Pan, J. Wang, and A. V. Vasilakos, "Machine learning on big data: Opportunities and challenges," *Neurocomputing*, vol. 237, pp. 350–361, may 2017. [Cited on page 27]

[88] Y. Li, W. Jiang, L. Yang, and T. Wu, "On neural networks and learning systems for business computing," *Neurocomputing*, vol. 275, pp. 1150–1159, jan 2018. [Cited on page 27]

[89] R. Dubey, A. Gunasekaran, S. J. Childe, T. Papadopoulos, Z. Luo, S. F. Wamba, and D. Roubaud, "Can big data and predictive analytics improve social and environmental sustainability?," *Technological Forecasting and Social Change*, vol. 144, pp. 534–545, jul 2019. [Cited on page 27]

[90] A. L'Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, "Machine learning with big data: Challenges and approaches," *IEEE Access*, vol. 5, pp. 7776–7797, 2017. [Cited on page 27]

[91] B. Mahesh, "Machine learning algorithms - a review," *International Journal of Science and Research (IJSR)*, 01 2020. [Cited on page 27]

[92] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, "Power to the people: The role of humans in interactive machine learning," *AI Magazine*, vol. 35, pp. 105–120, dec 2014. [Cited on page 28]

[93] M. A. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, J. M. McCauley, M. J. Franklin, S. Shenker, and I. Stoica, "Fast and interactive analytics over hadoop data with spark," *login Usenix Mag.*, vol. 37, 2012. [Cited on page 28]

[94] T. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, "Machine learning in manufacturing: advantages, challenges, and applications," *Production & Manufacturing Research*, vol. 4, pp. 23–45, jan 2016. [Cited on page 28]

[95] K. H. Tan, Y. Zhan, G. Ji, F. Ye, and C. Chang, "Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph," *International Journal of Production Economics*, vol. 165, pp. 223–233, jul 2015. [Cited on page 28]

[96] L. Cai and Y. Zhu, "The challenges of data quality and data quality assessment in the big data era," *Data Science Journal*, vol. 14, p. 2, may 2015. [Cited on pages 28 and 29]

[97] D. Blanchard, *Supply Chain Management Best Practices*. Wiley & Sons, Limited, John, 2021. [Cited on page 28]

[98] G. Büchi, M. Cugno, and R. Castagnoli, "Smart factory performance and industry 4.0," *Technological Forecasting and Social Change*, vol. 150, p. 119790, jan 2020. [Cited on page 28]

[99] O. Osoba and W. Welser, *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. RAND Corporation, 2017. [Cited on page 28]

[100] A. Haug, F. Zachariassen, and D. van Liempd, "The costs of poor data quality," *Journal of Industrial Engineering and Management*, vol. 4, jul 2011. [Cited on page 28]

[101] M. Ghobakhloo, "The future of manufacturing industry: a strategic roadmap toward industry 4.0," *Journal of Manufacturing Technology Management*, vol. 29, pp. 910–936, jun 2018. [Cited on page 29]

[102] T. Zheng, M. Ardolino, A. Bacchetti, and M. Perona, "The applications of industry 4.0 technologies in manufacturing context: a systematic literature review," *International Journal of Production Research*, vol. 59, pp. 1922–1954, oct 2020. [Cited on pages 29 and 30]

[103] U. M. Dilberoglu, B. Gharehpapagh, U. Yaman, and M. Dolen, "The role of additive manufacturing in the era of industry 4.0," *Procedia Manufacturing*, vol. 11, pp. 545–554, 2017. [Cited on page 29]

[104] A. Majeed, Y. Zhang, S. Ren, J. Lv, T. Peng, S. Waqar, and E. Yin, "A big data-driven framework for sustainable and smart additive manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 67, p. 102026, feb 2021. [Cited on page 29]

[105] H. Kang, "The prevention and handling of the missing data," *Korean Journal of Anesthesiology*, vol. 64, no. 5, p. 402, 2013. [Cited on page 29]

[106] J. Osborne, *Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data.* SAGE Publications, Inc., 2013. [Cited on page 30]

[107] S. Jeble, R. Dubey, S. J. Childe, T. Papadopoulos, D. Roubaud, and A. Prakash, "Impact of big data and predictive analytics capability on supply chain sustainability," *The International Journal of Logistics Management*, vol. 29, pp. 513–538, may 2018. [Cited on page 30]

[108] T. Boone, R. Ganeshan, A. Jain, and N. R. Sanders, "Forecasting sales in the supply chain: Consumer analytics in the big data era," *International Journal of Forecasting*, vol. 35, pp. 170–180, jan 2019. [Cited on page 30]

[109] T. Papadopoulos, A. Gunasekaran, R. Dubey, N. Altay, S. J. Childe, and S. Fosso-Wamba, "The role of big data in explaining disaster resilience in supply chains for sustainability," *Journal of Cleaner Production*, vol. 142, pp. 1108–1118, jan 2017. [Cited on page 30]

[110] H. Fatorachian and H. Kazemi, "A critical investigation of industry 4.0 in manufacturing: theoretical operationalisation framework," *Production Planning & Control*, vol. 29, pp. 633–644, jan 2018. [Cited on page 30]

[111] R. Y. Zhong, X. Xu, E. Klotz, and S. T. Newman, "Intelligent manufacturing in the context of industry 4.0: A review," *Engineering*, vol. 3, pp. 616–630, oct 2017. [Cited on page 30]

[112] E. Oztemel and S. Gursev, "Literature review of industry 4.0 and related technologies," *Journal of Intelligent Manufacturing*, vol. 31, pp. 127–182, jul 2018. [Cited on page 30]

[113] S. Erevelles, N. Fukawa, and L. Swayne, "Big data consumer analytics and the transformation of marketing," *Journal of Business Research*, vol. 69, pp. 897–904, feb 2016. [Cited on page 30]

[114] C. F. Hofacker, E. C. Malthouse, and F. Sultan, "Big data and consumer behavior: imminent opportunities," *Journal of Consumer Marketing*, vol. 33, pp. 89–97, mar 2016. [Cited on page 30]

[115] L. Ardito, A. M. Petruzzelli, U. Panniello, and A. C. Garavelli, "Towards industry 4.0: Mapping digital technologies for supply chain management-marketing integration," *Business Process Management Journal*, vol. 25, pp. 323–346, jul 2018. [Cited on page 30]

[116] E. Manavalan and K. Jayakrishna, "A review of internet of things (iot) embedded sustainable supply chain for industry 4.0 requirements," *Computers & Industrial Engineering*, vol. 127, pp. 925–953, jan 2019. [Cited on page 30]

[117] Y. Wang, L. Kung, and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," *Technological Forecasting and Social Change*, vol. 126, pp. 3–13, jan 2018. [Cited on page 30]

[118] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems*, vol. 12, pp. 5–33, mar 2015. [Cited on page 30]

[119] C. Batini and M. Scannapieco, *Data and Information Quality Dimensions, Principles and Techniques*. Springer London, Limited, 2016. [Cited on page 31]

[120] D. Ardagna, C. Cappiello, W. Samá, and M. Vitali, "Context-aware data quality assessment for big data," *Future Generation Computer Systems*, vol. 89, pp. 548–562, dec 2018. [Cited on page 31]

[121] I. Blanquer and W. Meira, "EUBra-BIGSEA, a cloud-centric big data scientific research platform," in *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*, IEEE, jun 2018. [Cited on page 31]

[122] R. Addo-Tenkorang and P. T. Helo, "Big data applications in operations/supply-chain management: A literature review," *Computers & Industrial Engineering*, vol. 101, pp. 528–543, nov 2016. [Cited on page 31]

[123] M. Pournader, Y. Shi, S. Seuring, and S. L. Koh, "Blockchain applications in supply chains, transport and logistics: a systematic review of the literature," *International Journal of Production Research*, vol. 58, pp. 2063–2081, aug 2019. [Cited on page 31]

[124] S. Winkelhaus and E. H. Grosse, "Logistics 4.0: a systematic review towards a new logistics system," *International Journal of Production Research*, vol. 58, pp. 18–43, may 2019. [Cited on page 32]

[125] Y. Li, J. Dai, and L. Cui, "The impact of digital technologies on economic and environmental performance in the context of industry 4.0: A moderated mediation model," *International Journal of Production Economics*, vol. 229, p. 107777, nov 2020. [Cited on page 32]

[126] D. Ivanov and A. Dolgui, "A digital supply chain twin for managing the disruption risks and resilience in the era of industry 4.0," *Production Planning & Control*, vol. 32, pp. 775–788, may 2020. [Cited on page 32]

[127] J. Kees, C. Berry, S. Burton, and K. Sheehan, "An analysis of data quality: Professional panels, student subject pools, and amazon's mechanical turk," *Journal of Advertising*, vol. 46, pp. 141–155, jan 2017. [Cited on page 32]

[128] M. Chmielewski and S. C. Kucker, "An MTurk crisis? shifts in data quality and the impact on study results," *Social Psychological and Personality Science*, vol. 11, pp. 464–473, oct 2019. [Cited on page 32]

[129] S. S. Kamble, A. Gunasekaran, and S. A. Gawankar, "Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications," *International Journal of Production Economics*, vol. 219, pp. 179–194, jan 2020. [Cited on page 32]

[130] B. Esmaeilian, J. Sarkis, K. Lewis, and S. Behdad, "Blockchain for the future of sustainable supply chain management in industry 4.0," *Resources, Conservation and Recycling*, vol. 163, p. 105064, dec 2020. [Cited on page 32]

[131] A. Bryman, "Quantitative and qualitative research: further reflections on their integration," in *Mixing Methods: qualitative and quantitative research*, pp. 57–78, Routledge, jul 2017. [Cited on page 33]

[132] S. Kalra, V. Pathak, and B. Jena, "Qualitative research," *Perspectives in Clinical Research*, vol. 4, no. 3, p. 192, 2013. [Cited on page 33]

[133] P. Aspers and U. Corte, "What is qualitative in qualitative research," *Qualitative Sociology*, vol. 42, pp. 139–160, feb 2019. [Cited on page 33]

[134] T. L. Lash, M. P. Fox, and A. K. Fink, *Applying Quantitative Bias Analysis to Epidemiologic Data.* Springer New York, 2009. [Cited on page 33]

[135] W. K. Hoy, *Quantitative research in education.* SAGE Publications, 2010. [Cited on page 33]

[136] J. Smith and H. Noble, "Bias in research: Table 1," *Evidence Based Nursing*, vol. 17, pp. 100–101, aug 2014. [Cited on page 33]

[137] P. Galdas, "Revisiting bias in qualitative research," *International Journal of Qualitative Methods*, vol. 16, p. 160940691774899, dec 2017. [Cited on page 33]

[138] D. H. Grossoehme, "Overview of qualitative research," *Journal of Health Care Chaplaincy*, vol. 20, pp. 109–122, jun 2014. [Cited on page 33]

[139] M. Hennink, I. Hutter, and A. Bailey, *Qualitative Research Methods.* SAGE Publications, Limited, 2020. [Cited on page 33]

[140] J. McCambridge, J. Witton, and D. R. Elbourne, "Systematic review of the hawthorne effect: New concepts are needed to study research participation effects," *Journal of Clinical Epidemiology*, vol. 67, pp. 267–277, mar 2014. [Cited on page 33]

[141] K. W. R Legard, J Keegan, *Qualitative Research Practice.* Sage Publications Ltd, 2003. [Cited on page 34]

[142] L. Bardin, *L'analyse de contenu.* Collection 'Le Psychologue', Presses universitaires de France, 1977. [Cited on pages 35 and 39]

[143] H. Heath and S. Cowley, "Developing a grounded theory approach: a comparison of glaser and strauss," *International Journal of Nursing Studies*, vol. 41, pp. 141–150, feb 2004. [Cited on page 37]

[144] B. G. Glaser and A. L. Strauss, *Discovery of grounded theory: Strategies for qualitative research.* Routledge, 2017. [Cited on page 37]

[145] D. Walker and F. Myrick, "Grounded theory: An exploration of process and procedure," *Qualitative Health Research*, vol. 16, pp. 547–559, apr 2006. [Cited on page 37]

[146] G. Guest, A. Bunce, and L. Johnson, "How many interviews are enough?," *Field Methods*, vol. 18, pp. 59–82, feb 2006. [Cited on page 37]

[147] G. Guest, E. Namey, and M. Chen, "A simple method to assess and report thematic saturation in qualitative research," *PLOS ONE*, vol. 15, p. e0232076, may 2020. [Cited on page 37]