



# A Topic Modeling Approach to Determine Supply Chain Management Priorities Enabled by Digital Twin Technology

Hirata, Enna

Watanabe, Daisuke

Chalmoukis, Athanasios

Lambrou, Maria

---

## (Citation)

Sustainability, 16(9):3552

## (Issue Date)

2024-05

## (Resource Type)

journal article

## (Version)

Version of Record

## (Rights)

© 2024 by the authors. Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

## (URL)

<https://hdl.handle.net/20.500.14094/0100489658>



## Article

# A Topic Modeling Approach to Determine Supply Chain Management Priorities Enabled by Digital Twin Technology

Enna Hirata <sup>1,\*</sup> , Daisuke Watanabe <sup>2</sup> , Athanasios Chalmoukis <sup>3</sup> and Maria Lambrou <sup>3</sup>

<sup>1</sup> Graduate School of Maritime Sciences, Kobe University, Kobe 658-0022, Japan

<sup>2</sup> Department of Logistics and Information Engineering, Tokyo University of Marine Science and Technology, Etchujima, Tokyo 135-8533, Japan; daisuke@kaiyodai.ac.jp

<sup>3</sup> Department of Shipping, Trade and Transport, University of the Aegean, 811 00 Mitilini, Greece; a.chalmoukis@aegean.gr (A.C.); mlambrou@aegean.gr (M.L.)

\* Correspondence: enna.hirata@platinum.kobe-u.ac.jp

**Abstract:** Background: This paper examines scientific papers in the field of digital twins to explore the different areas of application in supply chains. Methods: Using a machine learning-based topic modeling approach, this study aims to provide insights into the key areas of supply chain management that benefit from digital twin capabilities. Results: The research findings highlight key priorities in the areas of infrastructure, construction, business, technology, manufacturing, blockchain, and agriculture, providing a comprehensive perspective. Conclusions: Our research findings confirm several recommendations. First, the machine learning-based model identifies new areas that are not addressed in the human review results. Second, while the human review results put more emphasis on practicality, such as management activities, processes, and methods, the machine learning results pay more attention to macro perspectives, such as infrastructure, technology, and business. Third, the machine learning-based model is able to extract more granular information; for example, it identifies core technologies beyond digital twins, including AI/reinforcement learning, picking robots, cybersecurity, 5G networks, the physical internet, additive manufacturing, and cloud manufacturing.

**Keywords:** artificial intelligence; topic modeling; digital twin; supply chain management; sustainability



**Citation:** Hirata, E.; Watanabe, D.; Chalmoukis, A.; Lambrou, M. A Topic Modeling Approach to Determine Supply Chain Management Priorities Enabled by Digital Twin Technology. *Sustainability* **2024**, *16*, 3552. <https://doi.org/10.3390/su16093552>

Received: 5 March 2024

Revised: 19 April 2024

Accepted: 19 April 2024

Published: 24 April 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Digital twins (DTs) in supply chains are at the forefront of innovation and represent an exciting opportunity for the industry. Digital twin applications in supply chains have been the focus of recent research. Some of the recent research topics include the analysis of production logistics of digital twins and their functional characteristics in various application processes such as material distribution, packaging, transporting, warehousing, and information handling [1]. Another area of research is the application of digital twins in urban logistics, where the use of digital twins can optimize urban traffic and improve the quality of living in cities [2]. In addition, the approach of digital supply chain digital twins (SCDTs) has been explored as a means to optimize urban logistics processes, but the implementation of holistic SCDTs has faced barriers [3]. Furthermore, research has examined the implementation of DTs in distribution networks, particularly in the context of logistics service providers collaborating with vendors in e-commerce activities [4,5]. Finally, there are studies discussing the application of DTs for supply chain resilience testing [6], key performance indicator setting [7], risk reduction [8], the management of circular supply chains [9], and digital spare parts supply [10].

DTs enable organizations to simulate and analyze the impact of different operational strategies on sustainability metrics by creating virtual replicas of physical supply chains. DT technology provides a comprehensive and integrated approach to modeling and managing supply chain processes, helping to improve efficiency, responsiveness, and overall

performance. This technology helps predict the impact of supply chain changes, assess the viability of green initiatives before they are implemented, and continuously monitor the environmental impact of operations. Researchers continue to explore innovative applications and advancements in this area.

The integration of natural language processing (NLP) with DT technology can provide a powerful toolset for identifying and prioritizing supply chain management (SCM) priorities. Motivated by this ideology, this study applies NLP technology to identify supply chain priorities in the adoption of DTs.

This study reviews more than 1700 academic papers published in journals indexed in Emerald, Springer, Science Direct, and Web of Science (WoS). The papers collected in PDF format are then converted to a computer-readable structural format and NLP techniques are applied to analyze the data. The goal of the analysis is to extract and summarize supply chain management priorities enabled by DT technology.

Our study reveals key priorities such as infrastructure, construction, business, technology, manufacturing, blockchain, and agriculture domains. Our research contributions can be summarized in three ways. First, this study identifies new areas, such as construction, blockchain, and agriculture domains, which are not addressed in the human review results. Second, the research results also suggest that the human review results put more emphasis on the practical aspects, such as management activities, processes, and methods, while the machine learning results pay more attention to macro perspectives, such as infrastructure, technology, and business. Third, the machine learning-based model was able to extract more granular information; for example, it identified the core technologies other than DTs, including artificial intelligence (AI), reinforcement learning, picking robots, cybersecurity, 5G networks, the physical internet (PI), additive manufacturing, and cloud manufacturing (CMFG).

The remainder of this paper is organized as follows: Section 2 presents a review of previous studies on the application of DTs to the supply chain and the application of topic models. Section 3 explains the collection and conversion of the data in this study. Section 4 details the machine learning model employed in this research study. Section 5 summarizes the key findings and visual representations of the research results. Section 6 discusses these key findings. Section 7 concludes the study with a discussion of the practical implications of this work and the potential avenues for future research.

## 2. Previous Studies

### 2.1. Application of the Digital Twin to the Supply Chain

Among the numerous and constantly updated conceptualizations, a DT can be defined as a collection of information and computational models that map physical objects, processes, and entire cyber-physical systems into a virtual environment. Physical entities, processes, and their virtual twin models exchange information bi-directionally, thus allowing us to monitor, predict, diagnose, simulate, and control the state and behavior of the cyber-physical systems under consideration, in our case supply chain management infrastructures and services [11–15]. Building and using a specific supply chain management DT provides certain benefits, primarily in terms of offering appropriate variations for the optimization of a green performance, being cost-effective and safe, and providing reliable operation of the system under consideration [16,17]. Specific goals related to the design, operation, and control of the cyber-physical system under consideration can be modeled, implemented, and vigorously validated.

Against this background, a common use of SCDTs is to (a) continuously monitor specific supply chain processes in real time and hence enable improved operational business decisions. In addition, SCDTs can (b) predict system behavior, i.e., using physics-based or rule-based models as well as machine learning techniques, and accordingly control future system behavior, henceforth improving performance aspects of specific supply chain processes. Supply chain management scenarios simulated with digital twins (c) provide a testing platform to validate different strategic, tactical, or operational scenarios and enable decision makers to select the most appropriate ones to optimize the system's business or

technical performance. Other use cases are used to support more focused aspects, such as (d) improving supply chain security and resilience, as DTs' functionalities provide a risk assessment of various supply chain management threats and vulnerabilities, such as detecting malicious actions on a system.

Among previous studies, Ivanov (2023) [18] is identified as the most comprehensive and culminating study that presents research areas of DTs in supply chain management. The author examines recent research efforts in SCDTs and categorizes them into the following relevant aspects:

1. **DT Technology:** The main focus of DT research is the exploration of various enabling technologies, including IoT, communication protocols, data analysis technology, machine learning algorithms, DT development platforms, and security solutions.
2. **SCM Processes:** Supply chain management processes and tasks that are analyzed using DTs include inventory management, manufacturing operations management, job scheduling, sustainability performance monitoring, and resiliency modeling.
3. **Management Activities:** These include end-to-end visibility and transparency of goods movement and information flow, improved communication and collaboration with suppliers, supply chain risk simulation and management, and improved decision making through real-time decision support systems.
4. **Modeling Methods:** These include specific techniques such as machine learning algorithms (reinforcement learning and neural networks), categorized as descriptive, predictive, and prescriptive methods.
5. **Human–AI Symbiosis:** Fewer studies have examined the interaction between human decision makers and DTs in the supply chain, and the implications of this interaction.
6. **The Scope of DT Application:** SCDTs have been studied at different stages and levels of the supply chain. The scope includes both intraorganizational and interorganizational perspectives, including production or warehouse processes.
7. **Business Models:** SCDTs enable new business and operational models, such as collaborative platforms, cloud supply chains, and supply chain as a service.

This study compares the above areas with the areas identified using a machine learning-based topic modeling algorithm. Our research question is as follows:

RQ: In the area of application of DT technology in supply chains, what are the similarities and differences in areas identified by a human review and a review performed by machine learning?

## 2.2. Topic Models

Topic modeling is a method of machine learning that is classified as unsupervised learning. It is designed to analyze text data using NLP. The objective is to identify the most significant topics within a collection of documents by detecting recurring word and phrase patterns, grouping related terms, and determining the most representative terms for each document. By studying linguistic patterns such as the frequency and co-occurrence of words, topic modeling aggregates content with common threads, highlighting the central topics addressed across the entire body of documents.

Several well-known algorithms are used in topic modeling, for instance Latent Dirichlet Allocation (LDA) [19], Latent Semantic Analysis (LSA) [20], Non-Negative Matrix Factorization (NMF) [21], Probabilistic Latent Semantic Analysis (pLSA) [22], Top2vec [23], and BERTopic [24].

Four different topic modeling algorithms were compared in a recent study by Egger and Yu [25] that focused on the analysis of Twitter data. They found BERTopic and NMF to be most effective, on top of Top2Vec and LDA. In this study, the BERTopic model was specifically chosen because of its high performance in the extraction of topics from the text corpus. Section 4 of our study provides detailed information about the BERTopic algorithm applied.

### 3. Data

This study uses a dataset that was obtained from the Emerald, Springer, Science direct, and Web of Science databases of scholarly papers. For this dataset, a search was performed on 24 October 2023, using the terms “digital twin” and “supply chain”. To build the corpus for analysis, we used a three-step process (Figure 1). A total of 2318 papers were found in the first stage, in which the preliminary search was conducted. In the second stage, after filtering out the irrelevant papers, a total of 1793 papers were obtained for analysis. In the third step, Python version 3.10.12 was used to extract information from the PDF documents and convert it into structured data in JSON format. The structured data were then used for the processing and training of the machine learning models.

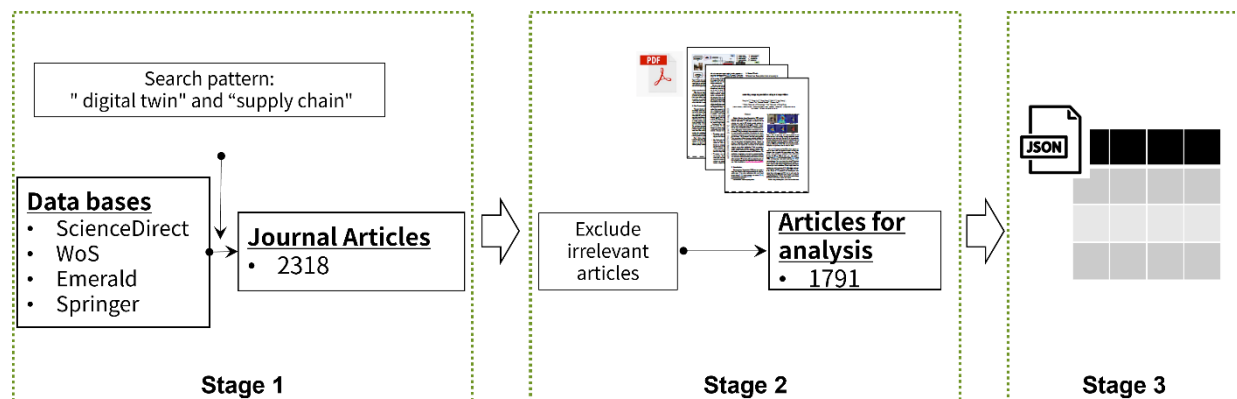


Figure 1. Data preparation.

To filter out irrelevant articles, we first tried to filter by keywords, which turned out to be inaccurate. We therefore decided to perform manual filtering through human reading, which resulted in the removal of 527 articles, from 2318 articles in stage one to 1791 articles in stage two.

A description of the articles selected in stage two is presented in Table 1.

Table 1. Description of data.

Database	No. of Articles	Contribution	No. of Words	Contribution
Emerald	274	15%	2,602,587	12%
ScienceDirect	952	53%	13,566,607	65%
Springer	362	20%	2,082,070	10%
WoS	203	11%	2,751,306	13%
Total	1791	100%	21,002,570	100%

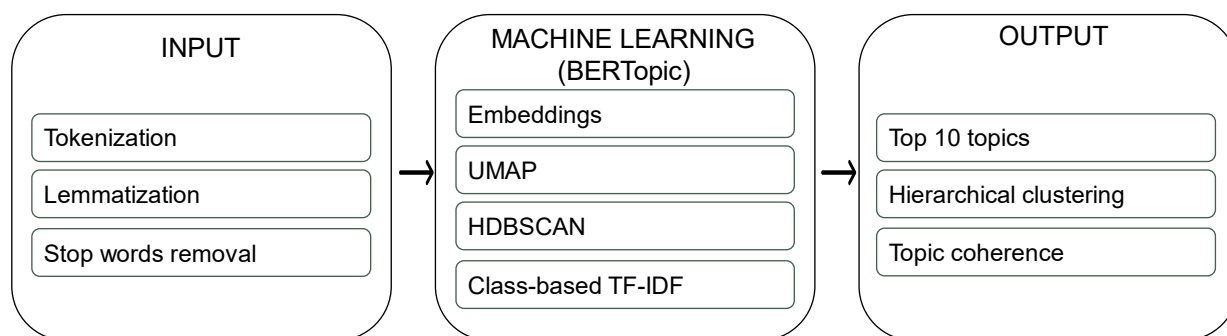
## 4. Methodology

### 4.1. Analysis Process

As shown in Figure 2, the analysis process is as follows. The input is prepared by data preprocessing including tokenization, lemmatization, and stop-word removal (Section 4.2). Machine learning is performed using the BERTopic model (Section 4.3). The output is presented and discussed in Section 5.

### 4.2. Preparation of Input Data

After collecting the data as detailed in the previous section, the next step is to preprocess the data in preparation for training the BERTopic model. The preprocessing consists of four steps using the Python package Gensim version 4.3.2 [26]. This package was chosen because of its open-source nature, efficiency, and capacity to handle large datasets.



**Figure 2.** Analysis process.

First, every word or punctuation mark is tokenized and treated as a separate unit. This facilitates the learning process of the model by splitting the text into several smaller segments. For example, “a digital twin is defined as a collection of information and computational models” is tokenized as “a”, “digital”, “twin”, “is”, “defined”, “as”, “a”, “collection”, “of”, “information”, “and”, “computational”, and “models”.

The second step is lemmatization. This involves identifying the root or base form of each token, known as a lemma. For example, the tokens “computer” and “computational” would both correspond to the lemma “compute”. The purpose of lemmatization is to prevent the creation of redundant elements that can result from different word forms.

Finally, in the third step, stop words are removed, which means commonly used words such as pronouns, determiners, and conjunctions. By removing stop words, the preprocessing step reduces noise in the data and improves the training process of the language model. In our example, stop-word removal removes the following tokens “a”, “is”, “as”, “of”, and “and”.

Natural Language Toolkit (NLTK) version 3.8.1 is used to perform tokenization and lemmatization. A composite stop-word list containing basic English stop words and stop words specific to this study such as “paper”, “et al.”, “journal”, “doi”, “vol”, and “pp” was used for stop-word removal.

#### 4.3. Machine Learning Using BERTopic Model

Building on the foundation of BERT, a language representation model proposed by Devlin et al. [27], BERTopic, is considered a deep learning-based approach to topic modeling. It uses a fine-tuning technique and is characterized by its pretraining strategy in NLP, which allows for the extraction of complex semantic information within sentences. In 2022, Grootendorst [24] presented an innovative solution that integrates transformer-based models with a class-based term frequency–inverse document frequency (TF-IDF) categorization to produce coherent and significant groupings, which ensures that important topic-describing keywords are maintained throughout the computation. Our research proposes a training procedure for BERTopic using a matrix-based text corpus. It outlines a five-step process for the topic modeling technique, which is elaborated in the next subsections.

##### 4.3.1. Embedding

The initial phase of the model involves transforming the corpora of text into structural data, i.e., numerical forms known as document and word embeddings. An embedding translates natural language into a computer-friendly numerical format. BERTopic performs this task by converting the input data into a mathematical representation using a sentence transformer [28]. While there are various methods for doing this, the model employs sentence transformers, specifically the “all-MiniLM-L6-v2”, which was adopted for its capacity to detect morphological similarities between documents.



#### 4.3.2. Uniform Manifold Approximation and Projection

Uniform Manifold Approximation and Projection (UMAP) is a technique used to reduce dimensionality when creating visual representations. It excels in the area of non-linear dimension reduction and is renowned for its effectiveness [29]. This is accomplished by transferring data from a high to low dimensional space while preserving the original dataset's intricate topological patterns.

#### 4.3.3. Hierarchical Density-Based Spatial Clustering of Applications with Noise

Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) is a statistical clustering approach introduced by Campello et al. [30]. This algorithm adopts a denseness-based clustering scheme, which permits it to bypass rigid delineations among groups. As a non-parametric methodology, it is specifically designed to detect the inherent high hierarchical pattern of clustering in the data by detecting regions densely populated by data points.

In the context of the BERTopic framework, HDBSCAN offers a clear advantage in its capability to detect clusters that do not adhere to traditional geometrical patterns. Instead, it focuses on areas where data density is significantly higher than in neighboring regions. This feature makes HDBSCAN highly valuable for identifying clusters of different densities.

#### 4.3.4. Class-Based Term Frequency–Inverse Document Frequency

The class-based term frequency–inverse document frequency (C-TF-IDF) is a refined form of the traditional TF-IDF method. It is specifically designed to detect clusters of topics by emphasizing the most important words within each cluster. As opposed to the standard TF-IDF approach, C-TF-IDF aggregates all relevant documents on a given topic into a unified document.

In every identified cluster or topic, denoted as 'c', the occurrence of a particular word 'x' is calculated and further refined through L1 normalization. This modification of C-TF-IDF allows for a more focused and more accurate representation of the meaningful words within each cluster or topic.

The calculation of the frequency of word 'x' in relation to topic 'c' is as follows:

$$W_{x,c} = \|tf_{x,c}\| \times \log\left(1 + \frac{A}{f_x}\right) \quad (1)$$

where  $tf_{x,c}$  stands for the occurrences of word  $x$  in class  $c$ ,  $f_x$  stands for the occurrences of word  $x$  in all classes, and  $A$  represents the mean number of words by class. The TF-IDF formula then takes the TF of a word and multiplies it by its IDF, which is derived by summing the logarithm of 1 and the mean score of words per class,  $A$ , divided by the score of word  $x$  in all classes.

This customized TF-IDF metric is constructed to extract distinct features of topics from document clusters by attaching a distinct tag to every topic. This version evaluates the contextually relevant meaning of words according to their relevance to a specific topic, as opposed to the standard TF-IDF, which assesses the meaning of words within individual documents. The method closely captures the occurrence of words across clusters, thus simplifying the generation of distinct distributions of words linked to each topic for the clusters of documents.

#### 4.3.5. Fine-Tuning

In the field of machine learning, the term fine-tuning refers to the process of optimizing a previously trained model that has already learned from a large-scale dataset in order to adapt it to a new (usually smaller and more specialized) dataset. The aim is to make the model suitable for a specific task or to improve its performance on a dataset that may slightly differ from the original training data. This procedure is based on the assumptions that the model can transfer the knowledge of features or the patterns that it learned from

the original data to the new data and adjust its parameters to meet the particular needs of the new one.

During the fine-tuning phase, the technique known as Maximal Marginal Relevance (MMR) [31] is applied. This technique is applied after C-TF-IDF representations have been created to identify a group of words that precisely represent a topic within a collection of documents. Though C-TF-IDF is effective in generating accurate topic representations, further refinement using MMR is necessary to ensure that these topics precisely represent the state of current discourse. MMR scores the similarity between single-word embeddings and the entire topic embedding by reducing the overlap of words across topics.

## 5. Experimental Findings

The results of the experiment have been obtained using the parameters listed in Table 2. A summary of these results is presented in the following subsections.

**Table 2.** The parameters of the experiment.

Parameter	Explanation	Utility
embedding model	The original BERTopic model was used for fine-tuning.	all-MiniLM-L6-v2
HDBSCAN	Density clustering algorithm using the excess of mass (EOM) method to select clusters.	min_cluster_size=5, and cluster_selection_method='eom'
UMAP	Dimensionality reduction model. 'n_neighbors' affects UMAP's compromise between local and global structure preservation, whereas 'n_components' is the desired dimensionality of the reduced embedding domain. The degree to which UMAP can group data items near one another is governed by 'min_dist'. Similarity computations are performed using the cosine distance.	n_neighbors=10, n_components=5, min_dist = 0.1, and metric='cosine'
Diversity	Evaluation of the variety of the selected terms and key phrases. Diversity utility ranges from 0 to 1, where 0 is minimum and 1 is maximum.	0.1

### 5.1. Top 15 Topics

Table 3 contains a summary of the top 15 topics. In addition to the keywords DT and supply chain that we used for the search, the most important key topics are identified as industry manufacturing data, digital business innovation, blockchain, computer science, bim, the physical internet (PI), AI and machine learning risk, the circular economy, cyber security, and cold chains.

**Table 3.** Overview of the information on the top 15 topics.

Topic	Topic Label	Count
1	0_industry_manufacturing_data	345
2	1_twin_digital_digital twin	176
3	2_supply_supply chain_chain	108
4	3_digital_business_innovation	71
5	4_blockchain_blockchain technology_chain	49
6	5_manufacturing_industry_computer science	36
7	6_bim_construction_building	35
8	7_construction_blockchain_contracts	25
9	8_logistics_pi_pl	24
10	9_ai_ml_risk	24
11	10_circular_recycling_circular economy	23
12	11_security_cyber_attacks	22
13	12_computer science_computer science	19
14	13_fruit_temperature_cold	18
15	14_ai_innovation_2020	17



The first topic relates to the use of data in the manufacturing industry. It may include the collection, analysis, and use of data to improve processes, optimize efficiency, and make informed decisions within the manufacturing sector.

The second topic involves the creation of a digital counterpart to monitor, analyze, and predict the behavior and performance of its real-world counterpart.

The third topic relates to the supply chain, which includes various stages, including sourcing, production, distribution, and logistics.

The fourth topic refers to the use of digital technologies and strategies to drive innovation within a company or organization. It involves the use of digital tools, processes, and platforms to transform business models, improve operations, enhance customer experiences, and create new opportunities.

The fifth topic is related to blockchain, which provides transparency, security, and immutability of the recorded data.

The sixth topic is about the intersection of computer science and the manufacturing industry. It may involve the application of computer science principles, algorithms, and technologies to improve manufacturing processes, automate tasks, and increase overall efficiency.

The seventh topic is about building information modeling (BIM), which provides a digital representation of the physical and functional characteristics of a building or infrastructure. It involves the creation and management of a 3D model that contains information about the design, construction, and operation of a structure. BIM is widely used in the construction industry to improve collaboration, reduce errors, and streamline the construction process.

The eighth topic relates to the use of blockchain technology in the construction industry, particularly in the context of contracts. Blockchain can enhance contract management by providing a secure and transparent platform for recording, verifying and executing construction contracts, ensuring trust and reducing disputes.

The ninth topic relates to logistics in the context of the physical internet (PI). The PI is a concept of a highly interconnected and collaborative logistics network that operates similarly to the digital internet. DT technology is used to enhance the operations and decision making within the PI framework.

The tenth topic is to explore the relationship between artificial intelligence (AI), machine learning (ML), and risk management. AI and ML technologies can be used to analyze large volumes of data and detect the patterns or anomalies that indicate potential threats or risks in various fields.

The eleventh topic is the circular economy, a business model that strives to reduce waste, increase resource efficiency, and drive sustainability.

The twelfth topic relates to cyber security and the threat of cyber-attacks. It may include a discussion of various security measures, strategies, and technologies used to protect computer systems, networks, and data from unauthorized access, breaches, and other malicious activities.

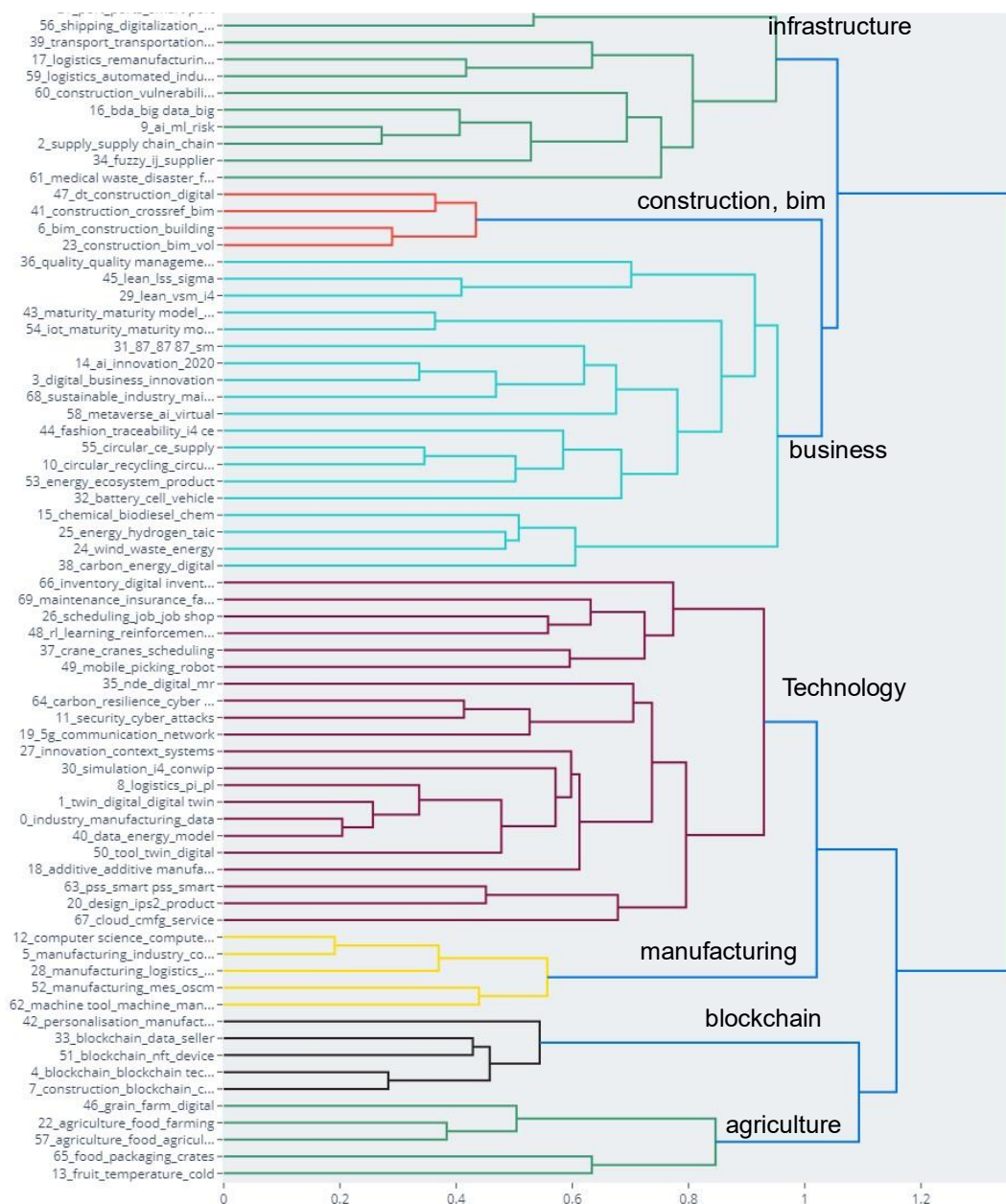
The thirteenth topic refers to the field of computer science itself. Computer science is the study of computation, algorithms, programming, and the design of computer systems. Its subject matter is broad, including databases, software development, artificial intelligence, and computer networks.

The fourteenth topic is related to the temperature requirements for fruit, especially the need for cold temperatures. In cold chain management, the application of DT technology helps to maintain freshness, extend shelf life, and prevent spoilage.

The fifteenth topic is focused on innovation and advances in AI. It may include a discussion of breakthroughs, developments, and applications of AI technologies that have emerged during the 2020 period.

### 5.2. Hierarchical Clusters

To gain deeper a perspective, hierarchical clustering was performed. This led to seven distinct clusters of topics, shown in different colors (Figure 3). The cumulative contribution rate was 99.1%, indicating that the seven clusters can explain almost the entire database. Details of the clustering results are discussed in Section 6.



**Figure 3.** Hierarchical clustering of topics.

### 5.3. Measurement of the Performance of the Model

We use a coherence score called C-umass [32] to assess the importance of our topics. This particular coherence score evaluates the frequency with which words occur together, relying solely on the documents used to train our model, without the need for external databases. It considers the frequency of co-occurrence of documents, checks the segmentation of adjacent documents, and uses the logarithm of the conditional probability as a verification measure. The scoring function used is the empirical conditional log-probability with smoothing to avoid calculating the logarithm of zero.

The C-umass coherence score is chosen for its fast computational performance, which is crucial for this study as it contains a large textual dataset.

A score close to 0 indicates that the topic is more coherent and understandable in a human sense. The overall coherence score of our model is  $-0.04$ , which means that it is very reliable.

## 6. Discussion

We then present the seven clusters with the top words in each cluster in Table 4, based on the clustering result in Figure 3.

**Table 4.** This is an overview of the topic clusters.

Cluster	Domains	Characteristics
1	Infrastructure domain	Smart ports Shipping digitalization Logistics automation
2	Construction domain	Construction Building information modeling (BIM) Lean production Maturity models Innovation
3	Business domain	Sustainability Circular economy Metaverse Energy sources and use (battery, wind, biodiesel, and carbon) AI/reinforcement learning Picking robots Cyber security
4	Technology domain	Digital twins 5G network Physical internet (PI) Additive manufacturing and cloud manufacturing (CMFG)
5	Manufacturing domain	Manufacturing logistics Operations and supply chain management (OSCM) Personalization
6	Blockchain domain	Blockchain data sellers Non-fungible token (NFT) Blockchain technology
7	Agriculture domain	Agriculture food Food packaging Cold chains

### 6.1. Infrastructure Domain

Cluster 1 can be broadly categorized under smart infrastructure in industry. This includes the integration of smart ports, shipping digitalization, logistics automation, and construction vulnerability. Smart ports use digital technologies such as Internet of Things (IoT) sensors, automated devices, robotics, big data analytics, and artificial intelligence to improve port operations and logistics efficiency. This enables the digitization of shipping through tools such as blockchain-based container tracking, predictive maintenance of assets, and autonomous ships/vehicles. Greater automation of logistics within and between ports helps reduce costs and errors. However, increased dependency on digital systems also introduces new design vulnerabilities. Port infrastructure and equipment require robust cybersecurity to protect sensitive data and operational technology networks from threats. Natural disasters pose challenges by potentially disrupting power, internet connectivity, and sensor infrastructure critical to automated functions.

### 6.2. Construction Domain

Cluster 2 is related to the construction domain, which includes construction and BIM. A DT takes the BIM model one step further by combining the physical and digital representations of a built asset. It links real-time sensor data from the physical structure to the virtual BIM model. This enables the monitoring and analysis of the physical structure's performance, operations, maintenance needs, and more over its lifetime.

Some potential concerns around BIM and digital twinning include the high initial costs and resource requirements. Developing detailed and accurate BIM models and integrating sensor networks require a substantial investment. Data quality, management, and security are also important challenges to ensure that virtual and physical assets are properly aligned. Standards and interoperability between multiple technology systems can be an issue. And there is a need to train the workforce to maximize the benefits of these advanced modeling and analysis approaches.

### 6.3. Business Domain

Cluster 3 contains lean production, maturity models, innovation, sustainability, circular economy, metaverse, and energy (battery, wind, biodiesel, and carbon). DT data drive continuous improvement by evaluating how changes in operations, technology, and energy use impact lean metrics and sustainability/maturity over time. Lean principles, maturity models, and the adoption of DTs help optimize supply chain efficiency and sustainability by reducing waste. Innovations around circular economy concepts such as battery recycling, biodiesel from waste, and carbon capture can be tested using DTs prior to implementation. Renewable energy sources such as wind and technologies such as biodiesel production facilities can be modeled as DTs to improve operations and return on investment, and support sustainability goals. Metaverse simulations allow for experimentation with supply chain scenarios under different sustainability and energy assumptions to identify risks, opportunities, and best practices.

Potential concerns include the cost of the initial investment, the need to ensure that DT data and metaverse scenarios are realistic, managing responsibilities around data governance and intellectual property, the willingness and ability of supply chain partners to collaborate digitally, and validating simulated results in the physical world. Standards and interoperability are key to realizing the full benefits.

### 6.4. Technology Domain

Cluster 4 consists of AI/reinforcement learning, picking robots, cyber security, DTs, 5G networks, the physical internet (PI), additive manufacturing, and cloud manufacturing (CMFG), which are core technologies for the adoption of DTs. AI techniques, in particular reinforcement learning, can be used to optimize picking robots, additive manufacturing, and other automations. DT simulations rely on machine learning-enabled AI. 5G networks and the PI enable real-time data exchange between physical objects and their DTs across manufacturing and logistics networks. Picking robots generate operational data that feed back into DT simulations to improve performance through reinforcement learning. Additive manufacturing and cloud/digital manufacturing use real-time data from physical processes, machines, and inventory for on-demand production optimized by AI. Cybersecurity is critical to protect connected physical assets, manufacturing equipment, operational data/IP, and DT simulations from threats in 5G, cloud, and distributed manufacturing environments.

It is interesting to note that the emerging concept of the PI is identified as a key technology for the adoption of DTs in the supply chain. The PI is a concept of logistics networks connected by standard modular containers and interfaces that enable the seamless transfer of physical and information objects. With PI networks facilitating the synchronized movement of both physical objects and their virtual data shadows, it would be much easier for companies to develop accurate, collaborative DTs that span the full range of end-to-end operations. This could help lower the barriers to the adoption of DT technologies for

monitoring performance, optimizing processes, and improving resiliency across a complex global supply chain. The synergies between the PI and DTs are clearly strong. Further development of both concepts in parallel is likely to be important to fully realize the vision of smarter, more transparent, and sustainable logistics systems in the coming years.

Potential concerns include workforce issues, ensuring AI safety and accountability, data privacy/ownership, dependency on fast and secure connectivity, protecting intellectual property, and maintaining resilience to cyberattacks across complex digital supply networks. Standardization is important to facilitate interoperability.

#### *6.5. Manufacturing Domain*

Cluster 5 is related to manufacturing logistics, and operations and supply chain management (OSCM). DTs of manufacturing plants, warehouses, and equipment help optimize logistics operations by providing simulations of material/component/product flows. OSCM strategies such as just-in-time (JIT) delivery, demand forecasting, inventory management, etc., can be tested using DTs prior to implementation. Logistics performance data from transportation, facilities, etc., feed back into DTs to continuously improve simulations based on real-world operations. Supply chain partners can collaborate digitally using shared DTs to identify bottlenecks, synchronize schedules, and improve resilience.

DTs help reduce the cost of prototyping by enabling a virtual evaluation of various “what-if” scenarios related to new processes, technologies, or disruptions. However, there are potential concerns, including initial investment costs, interoperability between partners’ systems, data security and privacy, ensuring that simulations reflect reality, overdependency on simulations over empirical learning, and change management for digital transformation. Standards and governance models are key to overcoming these challenges.

#### *6.6. Blockchain Domain*

Cluster 6 is about personalization, blockchain data sellers, non-fungible tokens (NFTs), and blockchain technology. Personalized DTs, based on consumer behavior data from blockchain data providers, could enable more tailored product/service recommendations. Supply chain stakeholders such as manufacturers could tokenize real-world assets as NFTs representing provenance, ownership, and usage rights data stored on a blockchain.

NFTs linked to DTs provide traceability of raw materials and components, and transparency into working conditions/environmental impacts. Their adoption depends on strategic alignment among supply chain partners in terms of governance, standards, and incentive structures.

#### *6.7. Agriculture Domain*

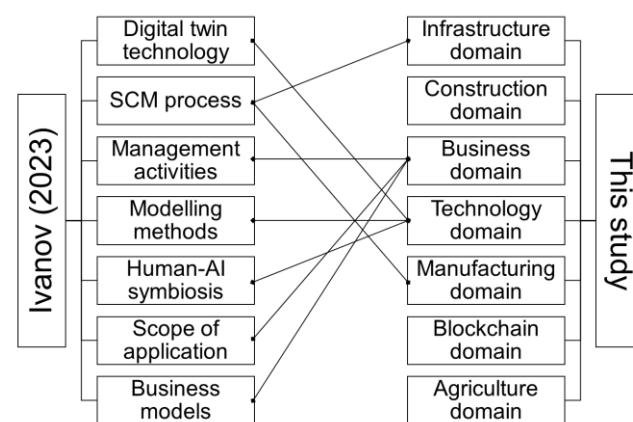
Cluster 7 includes agriculture food, food packaging, and cold chains. DTs of farms, food processing plants, warehouses, trucks, etc., can help optimize agriculture and food cold chain operations through simulations. Factors such as crop forecasting, equipment performance, and temperature/humidity control in packaging and transportation can be digitally modeled. Data from IoT sensors on food quality, shipments, etc., feed back to continuously improve DT simulations. Blockchain enables the traceability of food from farm to fork through digital records. In this context, partners can collaborate virtually to identify inefficiencies, synchronize schedules, and improve supply chain resilience to disruptions. Potential concerns include data quality and standards, technology/connectivity in remote locations, ensuring that simulations reflect real-world complexities, regulatory compliance, the initial investment in food safety when relying heavily on digital data, and change management for adoption. A proper governance structure would address privacy and liability issues in the case of failure.

Overall, while the technologies offer opportunities, concerns about privacy, standards, governance models, and change management would need to be addressed for successful DT-enabled supply chain transformations.

### 6.8. Comparison of Human Review and Machine Learning Results

In comparison to the human review results presented by Ivanov (2023) [18], which proposes seven areas, our research also identifies seven areas. A comparison is shown in Figure 4.

1. Compared with the human review result, this study generates new areas, such as construction, blockchain, and agriculture domains, which are not addressed in the human review results.
2. The human review results put more emphasis on practicality such as management activities, processes, and methods, while the machine learning results pay more attention to macro perspectives such as infrastructure, technology, and business.
3. By identifying the top keywords, the machine learning-based model was able to dig out more detailed information; for example, it identified the core technologies beyond DTs, including AI/reinforcement learning, picking robots, cybersecurity, 5G networks, the physical internet (PI), additive manufacturing, and cloud manufacturing (CMFG).



**Figure 4.** Comparison results of Ivanov (2023) [18] and this study.

## 7. Conclusions

In this study, advanced unsupervised machine learning techniques are used to analyze academic articles and identify key priorities related to the adoption of DTs in supply chains. The significance of this study is twofold. First, it extracts the top seven most important priorities to focus on for the adoption of DTs in supply chains. Secondly, it provides a different perspective on the adoption of DTs in supply chains from those generated by humans and, as such, aids in informed decision making and critical strategy building for the successful adoption of DTs in supply chains.

These research results show that existing research priorities are focused on SCM infrastructure and construction aspects, such as in business processes, models, and activities, as well as prioritizing digital technologies that enable DT capabilities. Mainly manufacturing processes and issues are examined, but more isolated research topics of special interest such as DT blockchains and DT technology for the agricultural domain are also identified.

Our research contributes in three ways. First, it identifies new areas, such as construction, blockchains, and agriculture domains, which are not addressed in the human review results. Second, the research results also suggest that the human review results put more emphasis on practicality, such as management activities, processes, and methods, while the machine learning results pay more attention to macro perspectives, such as infrastructure, technology, and business. Third, the machine learning-based model is able to extract more granular information; for example, it identifies the core technologies beyond DTs, including AI/reinforcement learning, picking robots, cybersecurity, 5G networks, the physical internet (PI), additive manufacturing, and cloud manufacturing (CMFG).

In addition, our study proves the utility of NLP methods in identifying and analyzing key priorities and tendencies in DT adoption in supply chains, thereby offering useful



insights for research and industry stakeholders. With our research results, we can observe that research clusters do not necessarily reveal a validated, evidence-based prioritization of research topics in terms of recognized research gaps or real-world industry problems. Henceforth, our research results reveal the need for more diverse research efforts to cover the following aspects: (a) various sectoral supply chain management problems investigated with DT capabilities, (b) diverse AI methods, in particular machine learning methods, for SCM predictive and prescriptive process monitoring and control, and (c) refined environmental policies, technologies, and related SCM processes investigated with DT capabilities.

In conclusion, our research helps society and practitioners to identify key sustainability issues, allowing sustainability initiatives to be prioritized and better tailored to stakeholder needs. This will eventually help foster collaborative decision making across the supply chain ecosystem. As such, leveraging these research implications can help supply chain companies become more resilient, sustainable, and responsive to the growing environmental and societal concerns that they are confronted with.

**Author Contributions:** Conceptualization, E.H. and D.W.; methodology, E.H.; software, E.H. and A.C.; validation, E.H., D.W. and M.L.; data curation, D.W. and A.C.; writing—original draft preparation, E.H., D.W., A.C. and M.L.; writing—review and editing, E.H., D.W., A.C. and M.L.; visualization, E.H.; funding acquisition, E.H. and D.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is supported by JSPS KAKENHI [Grant Number JP 23K04076, JP] and JSPS KAKENHI [Grant Number JP 21H01564].

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

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

## References

1. Zhu, Y.; Cheng, J.; Liu, Z.; Cheng, Q.; Zou, X.; Xu, H.; Wang, Y.; Tao, F. Production Logistics Digital Twins: Research Profiling, Application, Challenges and Opportunities. *Robot. Comput. Integr. Manuf.* **2023**, *84*, 102592. [\[CrossRef\]](#)
2. Abouelrous, A.; Bliet, L.; Zhang, Y. Digital Twin Applications in Urban Logistics: An Overview. *Urban Plan. Transp. Res.* **2023**, *11*, 2216768. [\[CrossRef\]](#)
3. Tasche, L.; Bähring, M.; Gerlach, B. Digital Supply Chain Twins in Urban Logistics System—Conception of an Integrative Platform. *Teh. Glas.* **2023**, *17*, 405–413. [\[CrossRef\]](#)
4. Kajba, M.; Jereb, B.; Cvahte Ojsteršek, T. Exploring Digital Twins in the Transport and Energy Fields: A Bibliometrics and Literature Review Approach. *Energies* **2023**, *16*, 3922. [\[CrossRef\]](#)
5. Kmiecik, M. Digital Twin as a Tool for Supporting Logistics Coordination in Distribution Networks. *Int. J. Supply Chain Manag.* **2023**, *12*, 1–6. [\[CrossRef\]](#)
6. Ivanov, D. Intelligent Digital Twin (iDT) for Supply Chain Stress-Testing, Resilience, and Viability. *Int. J. Prod. Econ.* **2023**, *263*, 108938. [\[CrossRef\]](#)
7. Marinagi, C.; Reklitis, P.; Trivellas, P.; Sakas, D. The Impact of Industry 4.0 Technologies on Key Performance Indicators for a Resilient Supply Chain 4.0. *Sustainability* **2023**, *15*, 5185. [\[CrossRef\]](#)
8. Astarita, V.; Guido, G.; Haghshenas, S.S.; Haghshenas, S.S. Risk Reduction in Transportation Systems: The Role of Digital Twins According to a Bibliometric-Based Literature Review. *Sustainability* **2024**, *16*, 3212. [\[CrossRef\]](#)
9. Preut, A.; Kopka, J.-P.; Clausen, U. Digital Twins for the Circular Economy. *Sustainability* **2021**, *13*, 10467. [\[CrossRef\]](#)
10. Peron, M. A Digital Twin-Enabled Digital Spare Parts Supply Chain. *Int. J. Prod. Res.* **2024**, 1–16. [\[CrossRef\]](#)
11. Sharma, A.; Kosasih, E.; Zhang, J.; Brintrup, A.; Calinescu, A. Digital Twins: State of the Art Theory and Practice, Challenges, and Open Research Questions. *J. Ind. Inf. Integr.* **2022**, *30*, 100383. [\[CrossRef\]](#)
12. Boyes, H.; Watson, T. Digital Twins: An Analysis Framework and Open Issues. *Comput. Ind.* **2022**, *143*, 103763. [\[CrossRef\]](#)
13. Bhandal, R.; Meriton, R.; Kavanagh, R.E.; Brown, A. The Application of Digital Twin Technology in Operations and Supply Chain Management: A Bibliometric Review. *Supply Chain Manag. Int. J.* **2022**, *27*, 182–206. [\[CrossRef\]](#)

14. Ivanov, D. Digital Supply Chain Management and Technology to Enhance Resilience by Building and Using End-to-End Visibility during the COVID-19 Pandemic. *IEEE Trans. Eng. Manag.* **2021**. [\[CrossRef\]](#)
15. Nguyen, T.; Duong, Q.H.; Van Nguyen, T.; Zhu, Y.; Zhou, L. Knowledge Mapping of Digital Twin and Physical Internet in Supply Chain Management: A Systematic Literature Review. *Int. J. Prod. Econ.* **2022**, *244*, 108381. [\[CrossRef\]](#)
16. Zhang, Z.; Guan, Z.; Gong, Y.; Luo, D.; Yue, L. Improved Multi-Fidelity Simulation-Based Optimisation: Application in a Digital Twin Shop Floor. *Int. J. Prod. Res.* **2022**, *60*, 1016–1035. [\[CrossRef\]](#)
17. Yan, Q.; Wang, H.; Wu, F. Digital Twin-Enabled Dynamic Scheduling with Preventive Maintenance Using a Double-Layer Q-Learning Algorithm. *Comput. Oper. Res.* **2022**, *144*, 105823. [\[CrossRef\]](#)
18. Ivanov, D. Conceptualisation of a 7-Element Digital Twin Framework in Supply Chain and Operations Management. *Int. J. Prod. Res.* **2023**, *62*, 2220–2232. [\[CrossRef\]](#)
19. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
20. Deerwester, S.; Dumais, S.T.; Furnas, G.W.; Landauer, T.K.; Harshman, R. Indexing by Latent Semantic Analysis. *J. Am. Soc. Inf. Sci.* **1990**, *41*, 391–407.
21. Lee, D.D.; Seung, H.S. Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature* **1999**, *401*, 788–791. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Hofmann, T. Probabilistic Latent Semantic Analysis. *arXiv* **2013**, arXiv:1301.6705v1.
23. Angelov, D. Top2Vec: Distributed Representations of Topics. *arXiv* **2020**, arXiv:2008.09470v1.
24. Grootendorst, M. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. *arXiv* **2022**, arXiv:2203.05794.
25. Egger, R.; Yu, J. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Front. Sociol.* **2022**, *7*, 886498. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Hirata, E.; Lambrou, M.; Watanabe, D. Blockchain Technology in Supply Chain Management: Insights from Machine Learning Algorithms. *Marit. Bus. Rev.* **2020**, *6*, 114–128. [\[CrossRef\]](#)
27. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**, arXiv:1810.04805.
28. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. *arXiv* **2019**, arXiv:1908.10084.
29. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2020**, arXiv:1802.03426v3.
30. Campello, R.J.G.B.; Moulavi, D.; Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. In *Advances in Knowledge Discovery and Data Mining*; Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2013; Volume 7819, pp. 160–172, ISBN 978-3-642-37455-5.
31. Carbonell, J.; Goldstein, J. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Melbourne, Australia, 24–28 August 1998; ACM: Melbourne Australia, 1998; pp. 335–336.
32. Röder, M.; Both, A.; Hinneburg, A. Exploring the Space of Topic Coherence Measures. In *Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, Shanghai, China, 2–6 February 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 399–408.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.