information

Article

Exploring the Features and Trends of Industrial Product E-Commerce in China Using Text-Mining Approaches

Zhaoyang Sun [1], Qi Zong [2], Yuxin Mao [2,*] and Gongxing Wu [2]

[1] Division of High-Tech Standardization, China National Institute of Standardization, Beijing 100191, China
[2] School of Management and E-Business, Zhejiang Gongshang University, Hangzhou 310018, China
* Correspondence: maoyuxin@zjgsu.edu.cn

Abstract: Industrial product e-commerce refers to the specific application of the e-commerce concept in industrial product transactions. It enables industrial enterprises to conduct transactions via Internet platforms and reduce circulation and operating costs. Industrial literature, such as policies, reports, and standards related to industrial product e-commerce, contains much crucial information. Through a systematical analysis of this information, we can explore and comprehend the development characteristics and trends of industrial product e-commerce. To this end, 18 policy documents, 10 industrial reports, and five standards are analyzed by employing text-mining methods. Firstly, natural language processing (NLP) technology is utilized to pre-process the text data related to industrial product commerce. Then, word frequency statistics and TF-IDF keyword extraction are performed, and the word frequency statistics are visually represented. Subsequently, the feature set is obtained by combining these processes with the manual screening method. The original text corpus is used as the training set by employing the skip-gram model in Word2Vec, and the feature words are transformed into word vectors in the multi-dimensional space. The K-means algorithm is used to cluster the feature words into groups. The latent Dirichlet allocation (LDA) method is then utilized to further group and discover the features. The text-mining results provide evidence for the development characteristics and trends of industrial product e-commerce in China.

Keywords: industrial product e-commerce; e-commerce platform; text mining; policy research; standard analysis

Citation: Sun, Z.; Zong, Q.; Mao, Y.; Wu, G. Exploring the Features and Trends of Industrial Product E-Commerce in China Using Text-Mining Approaches. *Information* **2024**, *15*, 712. https://doi.org/10.3390/info15110712

Academic Editor: Diego Reforgiato Recupero

Received: 25 September 2024
Revised: 3 November 2024
Accepted: 4 November 2024
Published: 6 November 2024

1. Introduction

Industrial products refer to products and services that are mainly used in industrial production or for maintaining the operation of an enterprise. These products are typically not directly sold to individual consumers but are procured and used by enterprises as means of production or maintenance, repair, and operation materials. According to the purpose of use by the purchaser and the original nature of the product, they are divided into two categories: non-production materials purchased for maintenance, repair, and operation, and items directly incorporated into the production process, namely, production materials. Industrial product e-commerce refers to the specific application of the e-commerce concept in industrial product trading and circulation. It entails the online display, sales, procurement, and related service activities of industrial products via Internet platforms [1]. In contrast to traditional offline transaction methods, industrial product e-commerce employs digital technology [2] to realize the online and automated processes of information exchange, transaction matching, contract signing, payment settlement, and logistics distribution between supply and demand parties. Industrial enterprises utilize e-commerce technologies, and both parties in transactions complete the entire process through real-time interaction, accessing market information and reducing implementation costs [3].

In recent years, the development momentum of the industrial Internet has been strongly influenced by progress in policy promotion, market demand, and digital technology. As a result, many countries and regions around the world attach great significance to the development of industrial product e-commerce. For instance, the EU has continuously strengthened its digital development strategy and issued a series of industrial policies to promote the implementation of the Industry 5.0 strategy and the comprehensive digital transformation of the industry in the region. With the vigorous promotion of the industrial Internet and the Industry 5.0 strategy by the EU and its member states, digital transformation has become the core driving force of the European industrial product market [4]. Meanwhile, Europe's leading industrial companies, such as Siemens, Dassault, and SAP, are leading industry innovation and transformation by building advanced digital service platforms. These platforms not only optimize the management efficiency of the supply chain but also promote the personalization and digitalization of products and services, inject high-value-added products and technical support into industrial e-commerce, and further catalyze the prosperity and diversity of the market. In China, the digital economy continuously accelerates the development of the B2B market for industrial products. To address the pain points of industrial development, the government has successively issued relevant policies to integrate the Internet economy with industrialized products. China's Ministry of Industry and Information Technology (MIIT) has also issued a series of policies related to industrial product e-commerce. These policies clearly state that industrial e-commerce should be further popularized and developed, and industrial product B2B platforms should be encouraged to jointly hold online trading activities for industrial products.

Although industrial product e-commerce has shown remarkable growth in recent years, the industry is still in transition from the traditional to the digital operation mode. At present, the scale efficiency and systematic construction of industrial product e-commerce are not yet fully mature, and there remains room for further optimization and expansion of development. First of all, regarding the operation mode, industrial product e-commerce must break its dependence on traditional offline transactions and address the problem of insufficient online informatization in the transaction process to improve the transparency and responsiveness of the overall supply chain. The information sharing, resource integration, and decision-making optimization capabilities of the digitization process will help reduce transaction costs and achieve more efficient matching of supply and demand, thus releasing market potential. Secondly, in terms of technical support, the booming development of industrial product e-commerce places higher demands on the technical foundation of the platform. With the widespread application of big data processing, intelligent algorithms, blockchain, and IoT technologies, industrial product e-commerce platforms demand a robust technical architecture and data security capabilities to achieve the storage, analysis, and privacy protection of massive data. This not only optimizes the user experience and safeguards transaction security but also provides real-time data feedback and market insights to support decision-making and business adjustment.

In addition, improving the service system is the key to enhancing the competitiveness of the industrial product e-commerce platform. In the traditional transaction mode, service support is often limited to basic transaction aggregation, while in the digital transformation, the construction and improvement of services such as logistics and distribution, quality testing, financial support, after-sales protection, and other value-added services significantly affect platform stickiness and customer satisfaction. The industrial product e-commerce platform requires a more comprehensive and professional service ecology to realize the integration of online and offline aspects to meet the diversified and personalized needs of enterprises and enhance the platform's comprehensive service capabilities. Finally, the policy environment and market standardization are also crucial to the development of industrial product e-commerce. Currently, the government continues to introduce policies that promote the development of industrial Internet and B2B e-commerce, aiming to build an open, transparent, and standardized trading environment by encouraging digital procurement, supporting the platform economy, and promoting supply chain collaboration.

However, the relative lag between market regulation and the standard system is a major obstacle to the development of the industry. The formulation and promotion of unified industry standards will help standardize platform operations, enhance market credibility, and promote the healthy development of the entire industry.

To promote the healthy development of industrial product e-commerce, we must rely not only on the guiding role of policy but also on the industry's own driving force. In-depth analysis of industrial product e-commerce-related policy documents, industry reports, standards, and other literature can help enterprises more accurately grasp the current situation and characteristics of the industry and better guide the development of the industry and related enterprises. Therefore, this study uses China's industrial product e-commerce-related literature as the data source and applies text-mining technology to conduct quantitative research and analysis to minimize the subjective bias often found in traditional manual analysis. Through text-mining technology, much information can be systematically processed and analyzed to extract valuable insights. After completing analytical steps, such as text preprocessing, keyword extraction, word vectorization, cluster analysis, and topic modeling, this study proposes specific countermeasure suggestions that promote the high-quality development of industrial product e-commerce.

This study aims to thoroughly analyze the core issues in China's industrial product e-commerce sector by exploring and validating an analytical framework based on text-mining techniques. This framework is designed to reveal the development characteristics, key trends, and challenges faced by the industry. The specific objectives include the following: (1) analyzing the content features of relevant policies, reports, and standards to extract characteristic terms and identify key industry directions; (2) uncovering the core themes of industrial product e-commerce from policies, reports, and standards, and analyzing the distribution and interrelations of these themes within documents to support industry policy formulation and corporate strategic planning; (3) identifying problems existing in industry policies and standards and proposing targeted countermeasures and recommendations. The ultimate research outcomes will help businesses better understand policy orientations and industry dynamics, provide the government with reliable empirical evidence and quantitative reference indicators, refine supportive policies for industrial product e-commerce, optimize the construction of standard systems, and promote the sustained and healthy development of China's industrial product e-commerce.

2. Related Works

Researchers have employed various theories to study the impact of e-commerce on industrial products. According to different research focuses, existing efforts can be divided into the following categories:

(1) Research on e-commerce and its role in the transformation of traditional industrial enterprises. Zhang et al. [5] proposed the concept of e-commerce embeddedness for integrating e-commerce and the manufacturing industry, and they analyzed the role of e-commerce in industrial research and innovation mechanism development. The results show that e-commerce embeddedness significantly affects R&D investment and manufacturing enterprises and reveal its role in the innovation and upgrading mechanism. Claycomb et al. [6] empirically tested different models using overall B2B e-commerce use as the dependent variable and industrial firms' innovation characteristics, environment, channel factors, and organizational structure as predictor variables. The researchers found that factors such as compatibility with existing systems, technological specialization, and information technology decision-making facilitated the overall use of B2B e-commerce in industrial firms, which in turn enhanced the value of the firm's performance.

(2) Research on obstacles to the development of the industrial product e-commerce model. Chen et al. [7] maintained that the key factor restricting the large-scale growth of industrial product e-commerce is its backward development strategy, such as the conflict between the e-commerce model and the industrial sales model and the lack

of professional technical and service support. Waithaka and Mnkandla [8] identified technical, security, cost, and environmental issues, along with the lack of computer knowledge, as obstacles to Kenya's manufacturing industry adopting the B2B market.

(3) Research on how e-commerce promotes industrialization. Tang and Wu [9] established a power model using the principle of system dynamics to discuss the various effects of e-commerce on the process of promoting industrial development; they established a causal loop diagram and a system flow diagram and analyzed the mechanism of e-commerce's impact on industrial development. Wang [10] explored how cloud computing and e-commerce affect industrial companies and industries in terms of technology architecture, service model, and industry chain.

(4) The impact of industrial product e-commerce on technological advancements in other industries. Al et al. [11] analyzed the impact of e-commerce on the digital transformation of the supply chain through a literature review and conceptual framework, especially in industrial organizations where e-commerce is seen as the embodiment of digital transformation. Sharma et al. [12] discussed e-commerce and digital transformation trends, opportunities, and challenges, highlighting how digital transformation is transforming modern business models through e-commerce and how emerging technologies, such as artificial intelligence, blockchain, and the Internet of Things, can play a key role in driving this transformation. Ren et al. [13] analyzed the application of deep generative models (DGMs) to large-scale generative models (LSGMs) for industrial time series generation and proposed a DGM-based AIGC framework, emphasizing that large models can act as drivers for industrial development and predicting and optimizing industrial processes.

(5) General research on using text mining in policy or other texts in other fields. Juventia et al. [14] used text mining to quantify countries' commitments to safeguarding and using agrobiodiversity. The study extracted and scored relevant sections of official documents, revealing varying levels of commitment among countries. Puri et al. [15] applied commonsense knowledge to the text mining of urban policy documents and social media postings. The approach uses reasoning based on commonsense knowledge to better account for pragmatics and semantics in the text, providing insights into public satisfaction and policy effectiveness. Tobback et al. [16] proposed an improved method for measuring economic policy uncertainty using text-mining techniques. The study compared traditional keyword-based methods with modality annotation and support vector machine (SVM) classification. Rao and Dey [17] discussed how text-mining techniques can assist in decision support for e-governance by retrieving and analyzing information from textual data sources. The study presents an integrated text-mining-based architecture to help policymakers discover associations between policies and citizens' opinions expressed in electronic public forums and blogs.

The existing literature reveals the impact of industrial product e-commerce on enterprise transformation, model innovation, and industrialization promotion, but there remain research gaps. First of all, although the existing studies provide a great deal of theoretical support in the integration mechanism, barriers, and paths to promoting the industrialization of e-commerce and the industrial products field, they mostly focus on theoretical discussions and lack specific quantitative analysis. From the particular perspective of text mining, systematic research on industrial product e-commerce is insufficient. Most of the research fails to fully integrate the industry's unique technological needs, market characteristics, and policy environment and is limited to a shallow analysis of patterns and barriers, making it difficult to provide specific solutions or guidance for the industry's development. In addition, industrial product e-commerce is less likely to adopt the quantitative method of text mining to conduct in-depth analyses of policy orientation, market demand, and technological development, which leads to a disconnect between the research and the actual development needs in the field and affects the application value of the theoretical results in practice.

Text mining is an increasingly prominent research tool in data-mining technology and aims to reveal hidden patterns and regularities in large-scale text data [18]. Policy documents and industry reports contain much information. Compared with traditional manual analyzing methods, text-mining technology can effectively deal with problems such as large amounts of text and strong subjectivity. It has more advantages than other data-mining techniques when analyzing the policies, reports, and standards related to industrial product e-commerce. Therefore, from the perspective of text mining, research related to the industrial product e-commerce industry contains much explorable space, and future research should dig deeper into the industry's needs.

3. Methodology

3.1. Data Collection

This study used industry policies, reports, and standards as research data because they present information about industry development from different dimensions. Policies are typically formulated and issued by governments at all levels. Thus, we downloaded the related policies mainly from government websites. The reports were collected from different industry research institutes or consulting companies. The standards mainly originated from various industry organizations and relevant government departments. After collection and screening, the research data comprised 18 policy documents from 2015 to 2023, 10 research reports, and 5 national standards from China. Some sample policies, reports, and standards are listed in Tables 1–3. All the texts are written in Chinese.

Table 1. Sample policies related to industrial product e-commerce.

No.	Policy Title	Issuing Department	Date
1	Guiding Opinions on Deepening the Integration and Development of Manufacturing and the Internet	State Council (PRC)	May 2016
2	Three-Year Action Plan for the Development of Industrial E-Commerce	MIIT of China	September 2017
3	Guiding Opinions on Promoting the Orderly Reopening of Industrial Communications Enterprises	MIIT of China	February 2020
4	Development Plan for the Deep Integration of Informatization and Industrialization under the 14th Five-Year Plan	MIIT of China	November 2021

Table 2. Sample reports related to industrial product e-commerce.

No.	Report Title	Research Organization	Date
1	China Manufacturing Industry Internet C2M E-Commerce Industry Research Report	iResearch	May 2019
2	Industrial E-Commerce White Paper	Department of Information Technology and Software Services, MIIT of China	July 2019
3	2020 China Industrial Products B2B Industry Research Report	36Krypton Research Institute	December 2020
4	Research on the Development and Investment Value of China's Industrial E-Commerce in 2020	CCID Consultants	December 2020
5	2022 China Industrial Products B2B Industry Research Report	iResearch	July 2022

Table 3. Sample standards related to industrial product e-commerce.

No.	Standard Title	Standard Number
1	E-Commerce Supplier Evaluation Criteria Quality Manufacturers	GB/T 30698-2014 [19]
2	Industrial Internet Platform Application Implementation Guide Part 2: Digital Management	GB/T 23031.2-2023 [20]
3	Industrial Internet Platform Selection Requirements	GB/T 42562-2023 [21]
4	Industrial Internet Platform Application Implementation Guide Part 5: Personalized Customization	GB/T 23031.5-2023 [22]
5	Industrial Internet Platform Application Implementation Guide Part 6: Service Extension	GB/T 23031.6-2023 [23]

3.2. Text-Mining Framework for Literature Analysis

Text-mining approaches have been widely used in analyzing industrial literature to uncover underlying patterns. In this regard, the design of an effective text-mining framework becomes essential. The design of the framework entails the planning and organization of the entire task, and a well-designed framework can enhance efficiency and ensure the achievement of text-mining goals. In this study, the literature relating to industrial product e-commerce was analyzed in detail using the framework illustrated in Figure 1. In this study, Python 3 libraries and tools designed for Chinese text processing were used for word splitting, customized stop-word lists used for preprocessing, and appropriately adapted libraries, such as sk-learn 1.2.1 and Gensim 3.0.0, were utilized for text mining. All the adopted techniques and methods were optimized for the characteristics of Chinese text to ensure that they could effectively support the subsequent data analysis work. The framework mainly comprises the following modules.

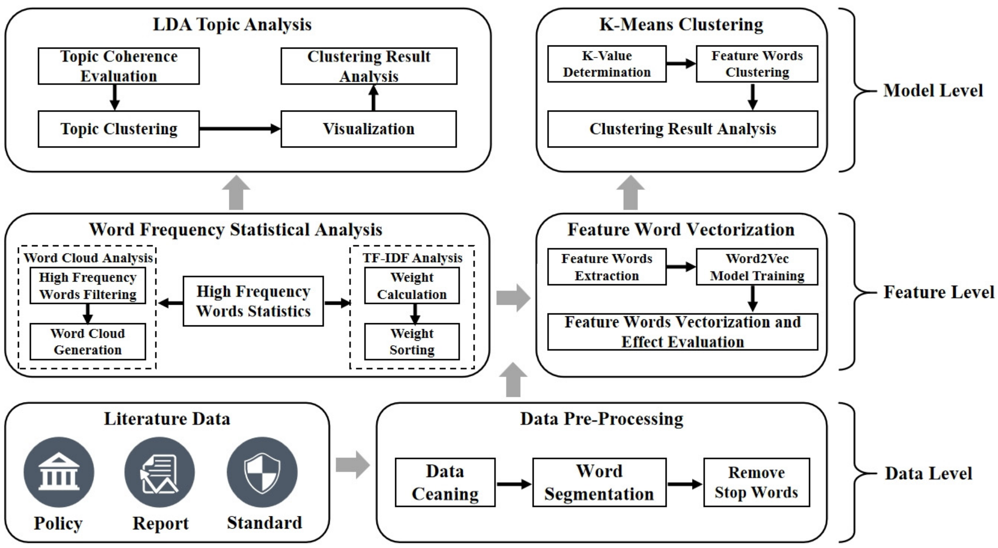

Figure 1. Framework of text mining in industrial product e-commerce.

(1) Data Pre-Processing: Data pre-processing is an important module for subsequent data analysis, which includes two key tasks: data cleaning and word segmentation. Firstly, we deal with the data in different formats from different sources and convert them into a unified format. Then, we split the text data into independent words and use the Jieba library [24] to segment the text data into separate words. Jieba is a popular Chinese text processing library in Python, specializing in Chinese word segmentation. Its main purpose is to solve the NLP challenges in Chinese text owing to the lack of clear word boundaries. The Jieba library identifies lexical units in a sentence by using a combination of statistical modeling and dictionary matching to accurately slice consecutive sequences of Chinese

characters into meaningful words. Subsequently, we optimize and adjust the segmentation results via stop-word filtering, retaining only key information. We can add new vocabulary words through a custom dictionary to maintain word integrity.

(2) Word Frequency Statistical Analysis: We obtain word frequency statistics to determine the keywords with high frequency in the text corpus, which helps to reduce data dimensions and thus alleviates the burden of subsequent model training. In this study, the TF-IDF [25] method is used to calculate the weights of the feature words in the text corpus, and these feature words are ranked according to the magnitude of their weights. The words with a TF-IDF value higher than a specified threshold are selected as the final feature words. This step significantly reduces the dimensions of the text model, providing a suitable model foundation for subsequent semantic calculations. Moreover, we can build a word cloud according to the word frequency statistics to further highlight high-frequency words.

(3) Feature Word Vectorization: By learning the distributed representation of words in the context, words with similar semantics are kept close to each other in the vector space. Word2Vec is a word embedding technique used to convert words in text into vector representations, which has advantages such as simple models, fast training speed, and the ability to effectively express similarities and analogical relationships between different words. Two possible models, Skip-Gram [26] and CBOW, can be used to train the Word2Vec model. Selecting the Skip-Gram model for word vectorization training results in more predictions, but through optimization of multiple parameters, the final word vector obtained is more accurate. Therefore, the Skip-Gram model was chosen for the vectorization training of feature words in this study.

(4) K-Means Clustering: The K-means algorithm [27] is a simple and effective unsupervised learning method that can quickly divide documents into a predefined number of clusters, thus helping to identify collections of documents with similar topics. The main advantage of K-means over other complex clustering algorithms, such as hierarchical or spectral clustering, is that it is simple to implement and computationally efficient. Especially when dealing with datasets, the iterative optimization process of K-means allows it to converge to a local optimum solution relatively quickly. Although it is sensitive to initial conditions and assumes spherical clusters, these limitations tend to become less significant in text analysis when the data are transformed by TF-IDF or word vectors.

First, the appropriate k value is determined using the profile coefficient, excellent samples are selected through the individual profile coefficients of the data objects, and the initial clustering center is adaptively selected for K-means clustering that optimizes the selection of initial clustering centers. The calculation formula of the profile coefficients is as follows:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{1}$$

The calculated sample $a(i)$ is the average distance from sample i to other samples in the same cluster, and $b(i)$ is the average distance from sample i to other clusters. The value of the profile coefficient is between -1 and 1. The closer it is to 1, the better the cohesion and separation are. The closer it is to -1, the worse the cohesion and separation are. The average of the profile coefficients of all the data points is the total profile coefficient of the clustering result.

Second, each sample is assigned to the cluster to which the nearest clustering center belongs. Third, the clustering centers are updated by calculating the average value of all the samples in each cluster and using it as the new clustering center. Fourth, the second and third steps are repeated, and if the distance between the newly calculated clustering center and the original center is less than a set threshold, the clustering is considered to reach the desired result and the algorithm stops.

(5) LDA Topic Analysis: LDA [28] is a generative model for text topic modeling. It assumes that each document is generated by a mixture of multiple topics, and each topic is generated by a group of words. Compared to other topic models, such as non-negative matrix factorization (NMF) or deep-learning-based approaches, LDA provides a probabilistic

explanatory framework in revealing underlying topics in text data. LDA not only identifies keywords that appear in a document but also provides probability distributions of each topic in each document, which is particularly useful for understanding and interpreting text data. Although LDA is computationally complex, it generally outperforms models that rely on feature engineering or require large amounts of labeled data on textual data. This study constructs a topic model and performs topic clustering on the pre-processed text data. The coherence calculation method is used to determine the number of topics, and then the topic hotspots are identified. Finally, the results of the LDA model analysis are visualized.

Topic coherence measures are used to evaluate the quality of topics in the LDA model. A common topic coherence measure is the topic coherence score, the basic idea of which is to measure the co-occurrence of words in the topic. One of the formulas for a topic coherence score is based on mutual information (PMI):

$$\text{Coherence } (V) = \sum_{m=2}^{M} \sum_{l=1}^{m-1} \log \frac{D(v_m, v_l) + \epsilon}{D(v_l)} \tag{2}$$

Topic strength is also called topic popularity or topic attention. It can show the degree of attention paid to a specific topic within a specific time period and is a quantitative indicator for judging the attention allocation of a specific topic. The formula for calculating topic strength is as follows:

$$P_k = \frac{\sum_i^N \theta_{ki}}{N} \tag{3}$$

P_k represents the strength of the kth topic, N is the number of texts, and θ_{ki} represents the probability of the kth topic in the ith text.

4. Findings

Based on the text-mining framework for industrial product e-commerce, we explored the development features and trends of industrial product e-commerce in depth.

4.1. Keyword Frequency Analysis

Keyword analysis plays an important role in revealing the core content of industrial literature. In this study, the Jieba library was used to perform the word-splitting process of the text data to obtain the keywords and their word frequency distribution. Subsequently, the keyword frequencies of all the documents were summarized and calculated, and the word segmentation results were ranked according to their frequency to generate a word cloud map of the industrial product e-commerce industry, as shown in Figure 2. Since the samples were industry support policies and market research reports, the document set after word separation contained high-frequency words, such as pronouns, quantifiers, convergent verbs, etc., which were not helpful for text characterization, so these words were eliminated. Finally, the top 20 effective high-frequency words were organized, as shown in Table 4.

The terms “industry”, “enterprise”, “platform”, “service”, and “e-commerce” had the highest frequency. This suggests that industrial product e-commerce and platforms are important engines for promoting the innovation and development of industrial enterprises. By supporting the key cultivation platforms of industrial product e-commerce, the government and the industry have pushed enterprises to accelerate platform construction. The frequency of terms “digitalization”, “data”, and “technology” arranged from high to low indicates that the industry is increasingly focused on achieving cost reduction and efficiency improvement in digital transformation to build a sound service system.

Figure 2. Industrial product e-commerce word cloud.

Table 4. Frequency of top 20 keywords.

Keyword	Word Frequency	Keyword	Word Frequency
industry	2542	data	728
enterprise	2068	ability	702
platform	1908	industry	695
service	1468	management	666
e-commerce	1198	production	648
Internet	1189	demand	633
development	1119	digitalization	624
product	897	B2B	610
procurement	841	technology	603
industrial products	758	manufacturing	568

To further show the results of the keyword frequency analysis, we generated a word cloud (see Figure 2) based on the keywords with high frequency. In the word cloud, keywords with larger font sizes appear more frequently in the documents. The keywords in the word cloud all appear in Chinese in the original texts.

Through the word cloud, we observed the prominence of “industry”, “enterprise”, “platform”, “service”, “e-commerce”, etc. The high frequency of these words reflects the core concerns in the field of industrial product e-commerce. In particular, the prominence of the terms “industry” and “enterprise” indicates the key role of industrial product e-commerce in promoting the development of industrial enterprises and industrial upgrading. Meanwhile, the high frequency of “platform” and “service” reveals that building an efficient and personalized e-commerce platform and providing high-quality services are the key factors for the success of industrial product e-commerce. In addition, the concentrated appearance of terms such as “digitalization”, “supply chain”, and “Internet” further emphasizes the importance of digital transformation and supply chain management in the development of industrial product e-commerce.

4.2. Feature Extraction and Vectorization

In the process of feature word extraction, we set the threshold value to 0.1, and only feature words with a TF-IDF value greater than 0.1 were selected. Table 5 shows some feature words with high TF-IDF values. After performing feature word extraction, we further trained the model using Word2Vec to obtain word vectors. Based on the feature word extraction, the model was further trained using Word2Vec to obtain the word vectors, setting the model parameters, as shown in Table 6. By processing the pre-processed text corpus, a corpus word list was obtained in which each word corresponded to a 200-dimensional space vector. The industrial product e-commerce feature words corresponded to 17,200-dimensional word vectors in this word list.

Table 5. Feature word screening and corresponding TF-IDF values.

Characteristic Word	TF-IDF Value
industry	0.404026256
enterprise	0.347823617
platform	0.283990818
e-commerce	0.222205136
development	0.205269903
service	0.203222787

Table 6. Model parameter settings for word vectorization.

Parameters	Instruction	Set Value
size	The length of the word vector	5
window	Context window length	200
skip-gram	Skip-gram modeling or not	1

The effectiveness of the trained word vectors was evaluated by observing the similarity between words and the list of related words for a single word. The four word pairs extracted from policy documents were "digitalization" and "technology", "manufacturing" and "production", "material" and "platform", and "user" and "Internet". The similarity of each word pair was calculated, and the results are shown in Table 7. Observing the word pair similarity results, we found that the processing results were consistent with daily human cognition, indicating that the Word2Vec training model could generate reasonable and effective word vectors. Therefore, the same operation could be conducted for other feature words in the corpus to verify the effectiveness of the word vectors obtained by training.

Table 7. Similarity of related word pairs.

Pair of Words	Similarity
<Digitalization, Technology>	0.83570686
<Manufacturing, Production>	0.91326806
<Material, Platform>	0.33396235
<User, Internet>	0. 26313045

4.3. K-Means Clustering

The silhouette coefficient is a metric employed to evaluate the quality of clustering, integrating the tightness within clusters and the separation between clusters. After obtaining the feature vectors, the silhouette coefficients were calculated for different values of k. By observing the trend of silhouette coefficients with k values in Figure 3, we found that the silhouette coefficients decreased gradually and fluctuated in the range of 0.06 to 0.44 as the k value increased. As shown in Table 8, in the cases where k was equal to 3, 4, and 5, the corresponding silhouette coefficients were the three largest values among all the silhouette coefficients corresponding to different k values, at 0.4377, 0.2673, and 0.2922. Since a silhouette coefficient closer to 1 indicates a better clustering effect, we chose to set the value of k to 5, considering that division into 3 and 4 clusters was impractical.

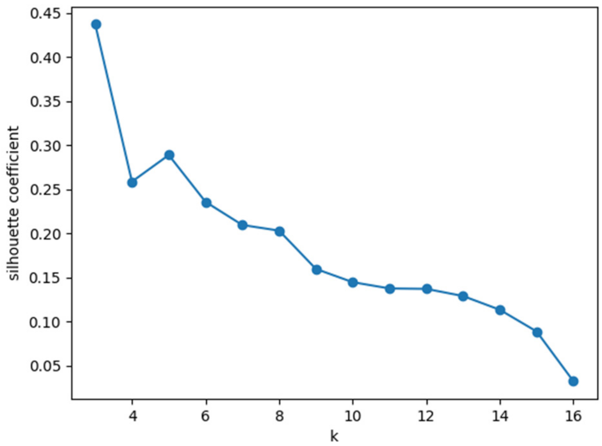

Figure 3. Silhouette coefficients corresponding to different values of k.

Table 8. Correspondence between different k values and silhouette coefficient.

K	Silhouette Coefficient	K	Silhouette Coefficient
3	0.4377	9	0.1403
4	0.2673	10	0.1427
5	0.2922	11	0.1322
6	0.2512	12	0.1411
7	0.1861	13	0.1266
8	0.1829	14	0.1162

After determining the value of k, the feature words were clustered using the K-means clustering algorithm. The details of the algorithm are given below. The feature screening and the corresponding categories are shown in Table 9 when k = 5. According to the clustering results in Table 9, each cluster represents a specific topic or related field within industrial product e-commerce. Cluster 1 is mainly the products and services provided by industrial product e-commerce to meet user needs. Cluster 2 mainly reflects the close connection of industrial product e-commerce to the development of the Internet and e-commerce. Cluster 3 mainly indicates that the B2B model is the most common operation mode in industrial product e-commerce. Cluster 4 mainly involves industrial enterprise development, platform construction, technology application, and digital transformation. Cluster 5 is mainly associated with the manufacturing production of industrial products.

In summary, these clusters represent aspects such as product, service, Internet, B2B, enterprise digitalization, technology construction, and manufacturing. It can be observed that the construction and policy development of China's industrial product e-commerce are gradually maturing. At present, industrial product e-commerce has advanced from the initial stage of platform construction to the stage of deep integration of industrial product trading, production, and services, while the policy focus has gradually shifted to ensuring the smooth flow of all aspects of the supply chain through industrial product e-commerce.

Table 9. Feature screening and corresponding categories.

Cluster	Featured Keywords
1	product, service
2	Internet, development, industry, e-commerce
3	B2B, industry, industrial
4	enterprise, platform, construction, technology, digitization, data
5	fabrication

From the perspective of the characteristics within the clusters, Cluster 2 and Cluster 3 involve the Internet, industry, and e-commerce, indicating that these two clusters are concerned with the interaction between the Internet and industrial product e-commerce in the development of e-commerce. National policies continuously promote the combination of the Internet and the industrial sector. This not only creates a favorable environment for the promotion and development of industrial product e-commerce but also promotes the practical application of the industrial Internet platform. The two promote each other to form a virtuous circle. Cluster 4 contains the most keywords, signaling the further extension of technology to the industrial side of industrial product e-commerce. The government and the industry are jointly committed to promoting the digital transformation of the supply chain. The transformation of enterprise management through digital technology has become an important direction in the industrial product e-commerce market competition. Cluster 1 reflects that the deepening of services is becoming the new focus of industrial product e-commerce. With the continuous development of industrial product e-commerce, the improvement of its service system has become an inevitable requirement. Due to the diversity and complexity of industrial products, both the supply and demand sides face problems. To solve these problems, the service system must be continuously improved. Cluster 5 is mainly associated with the manufacturing production of industrial products. Due to the relative maturity of manufacturing policies and industry research in relation to the industrial Internet and industrial product e-commerce, there is little need for extensive text narrative.

4.4. LDA Topic Modeling and Visualization

We drew a curve based on the number of optimal topics according to topic coherence, as shown in Figure 4. In general, higher topic coherence indicates stronger topic internal connections and higher interpretability.

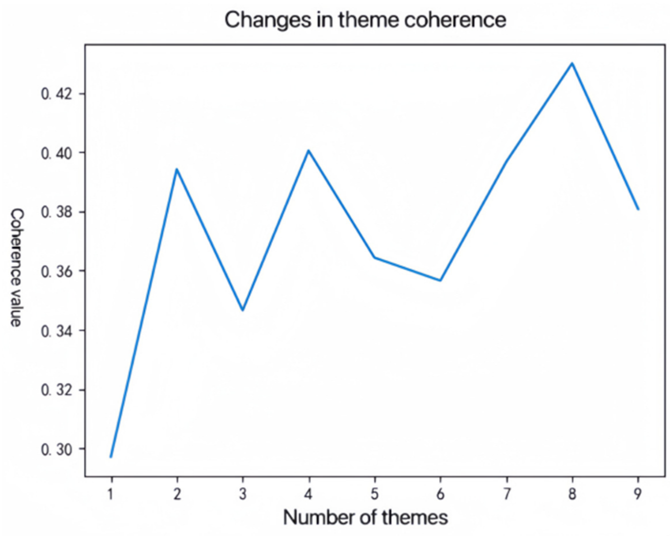

Figure 4. Topic coherence for LDA clustering.

Topic coherence was used to identify that the appropriate number of topics in this job was eight. Based on this, the number of model topics was set to eight, and the model was rerun to obtain the topic category to which each document most likely belonged. Through LDAvis 3.4.1, the visual documents of topic clustering, the visualization results, and the frequency distribution of feature words under each topic were obtained, as shown in Figure 5. The document results show the distribution and probability ratio of each of the eight topics. When a topic is selected, 30 representative hot keywords under the topic will emerge, which can be used to determine the topic content. The size of the circle intuitively reflects the significance and coverage of each topic in the entire text corpus, that is, the topic intensity. The larger the area of the circle, the higher the proportion of the topic in the entire corpus and the more significant its importance. The distance between the circles reveals the degree of correlation between topics; the closer the distance, the higher the correlation between topics. Figure 6 visualizes the word frequency distribution of Topic 1.

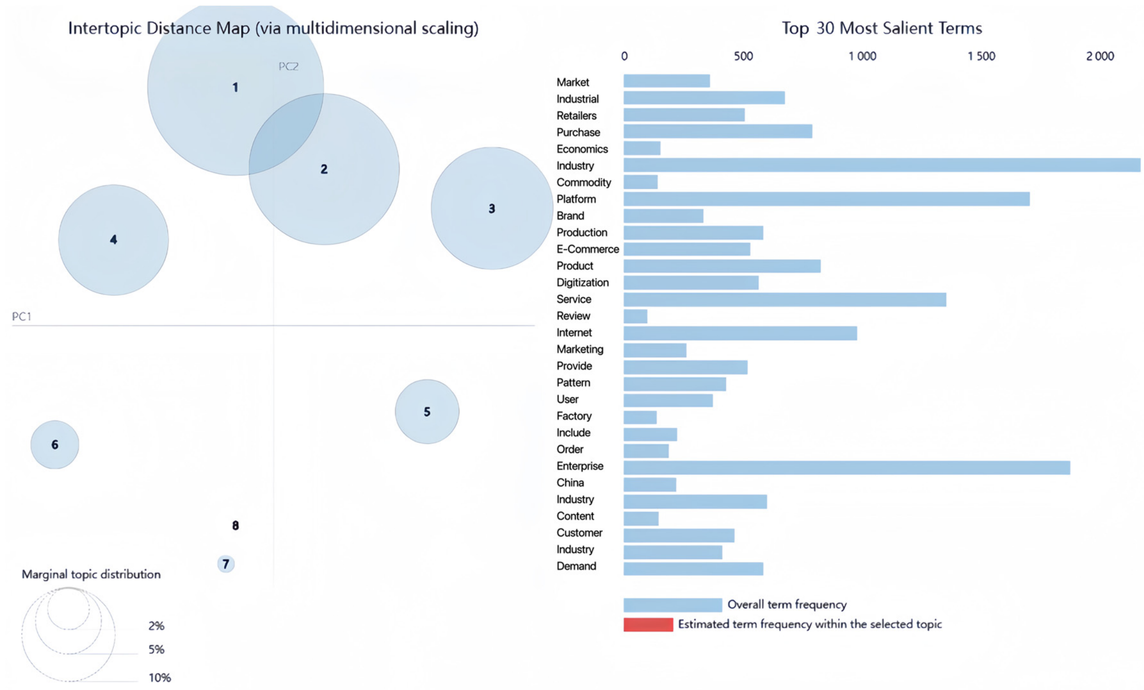

Figure 5. Frequency distribution of feature words under each topic.

The results of the LDA topic model analysis of the industrial product e-commerce-related texts are shown in Table 10. This study manually summarized the general titles related to each topic based on the text modeling results. According to the data on the proportion of topics, Topic 1 had the highest proportion, reaching 35.6%. These words cover core concepts, such as industry, platform, service, enterprise, and e-commerce, indicating that these are the hot topics in the industrial product e-commerce field in China. They demonstrate the importance of industrial product e-commerce platforms and related corporate management and market demand.

The other topics also have their own features. Topic 2 reflects the transformation and upgrading process of industrial product enterprises in the digital era and the changes in consumer behavior; Topic 3 emphasizes the impact of the macroeconomic environment on the circulation of industrial product e-commerce market and development trends in China; Topic 4 covers and emphasizes the changes triggered by Internet technology to traditional manufacturing and the exploration of new manufacturing models; Topic 5 reflects the role of industrial product e-commerce service providers in market competition and the impact of customer evaluation on the corporate image; Topic 6 explores the impact of changes in consumer behavior on the operating model of industrial product e-commerce platforms and the importance of data analysis; and the scope of Topics 7 and 8 was too small and thus was ignored.

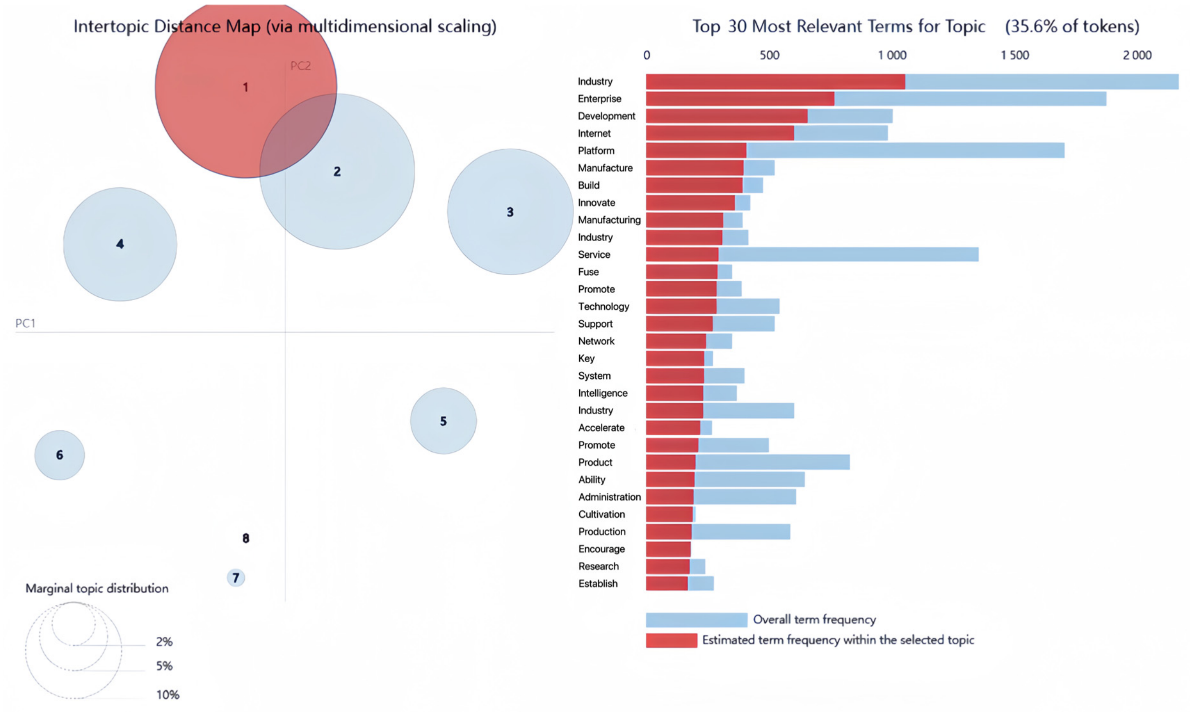

Figure 6. Frequency distribution of feature words under Topic 1.

Table 10. Results of the LDA topic model of industrial product e-commerce-related texts.

No.	Topic	Topic Feature Words (Top 8)	Topic Ratio (%)
1	Industrial product e-commerce platform	industry, platform, service, enterprise, e-commerce, demand, Internet, management	35.6
2	Digital transformation	enterprise, e-commerce, product, digitization, industry, industrial product, marketing, user	25.9
3	Market circulation	market, commodity, economy, circulation, entity, consumption, development, production and sales	17
4	Manufacturing innovation	industry, development, Internet, platform, manufacturing, construction, innovation, manufacturing industry	13.8
5	Service review	review, e-commerce, product, manufacturer, module, enterprise, provide, content	4.7
6	Consumption model	brand, factory, consumption, platform, model, order, demand, data	2.6

4.5. Comparative Analysis of Text-Mining Results

(1) Comparison of findings from the word frequency analysis and TF-IDF keyword extraction: In the text-mining process, the word frequency analysis indicates the frequency of use of all words, while TF-IDF filters out more discriminating keywords by considering their importance in the document collection. Both analysis methods highlighted the importance of "industry", "enterprise", "platform", "service", and "e-commerce" as keywords. "E-commerce" was a core topic in the industrial product e-commerce space, suggesting that these terms are key to understanding and analyzing the space. The combined use of the two methods provided a more comprehensive understanding of the focus and

structure of textual data, and the textual analyses complemented each other to provide a more balanced perspective. Enterprises or policymakers can formulate strategies based on these core keywords, especially in digital transformation, platform construction, and service optimization.

(2) Comparison of the findings from K-means clustering and LDA topic modeling: The K-means clustering method focuses on the distance between word vectors and tends to identify dense regions in the word vector space; it is therefore suitable for rough classification. In contrast, LDA topic modeling focuses on the probability distributions of words under different topics, providing a more nuanced and diverse topic delineation. This detailed delineation revealed deeper structures and relationships in textual data, helping companies or policymakers to identify specific topics to target their strategies. For example, LDA can identify segmented topics, such as "service reviews" and "consumption patterns", which can support the development of fine-grained service improvement initiatives and marketing strategies. Despite the differences in methodology and focus between K-means and LDA, they share the common goal of distilling the thematic structure hidden in industrial product e-commerce texts, which resulted in certain similarities. The overlap among the five clustering themes of K-means and the eight thematic models of LDA reflects their roles in identifying the key platform construction, B2B model, digital transformation, technology applications, and areas of consistency. Both methods captured the underlying structure of industrial product e-commerce texts. Therefore, using the two approaches together provided a more comprehensive understanding of the text data, suitable for both a generalized classification and in-depth analysis of the correlations among the themes, helping companies or policymakers to develop more precise strategies in the field of industrial product e-commerce.

5. Conclusions

This study used a text-mining methodology to explore the policies, reports, and standards related to industrial product e-commerce. The results of this study effectively address the initial objectives of this study and also provide empirical support and specific guiding suggestions for policy formulation, standard construction, and enterprise strategic planning in the field of industrial product e-commerce. To be more specific, the primary work and outcomes of this study can be summarized into the following three aspects:

(1) Analyzing text content features to identify key industry directions: This study found that keywords such as "platform" and "enterprise" rank high in policies, reports, and standards, indicating strong policy and industry support for platform innovation and the empowerment of industrial enterprises. This highlights that industrial product e-commerce platforms are at the core of driving the industrial product e-commerce ecosystem. Policies and the industry actively promote platform innovation and continuously empower industrial enterprises. In addition, industrial product e-commerce platforms are continuously enhancing their multifaceted functions, including supply chain finance and warehousing logistics services. While meeting the online sales needs of enterprises, these platforms also bring additional value-added benefits to industrial enterprises, providing favorable conditions for them to proactively integrate into new e-commerce business models.

(2) Mining and acquiring the core themes and structure of industrial product e-commerce literature: The research results reveal the focus of policies and the industry on the application of technology and the implementation of industrial product e-commerce. From the keyword extraction and clustering results, it was found that the attention of policies and the industry has shifted towards the integration of technology with the industry and applications, with a greater emphasis on the practical implementation of technologies related to industrial product e-commerce. In the traditional industrial supply chain channels, various roles, including agents, distributors, and retailers, are actively transforming to adapt to the trend of digitalization, fully leveraging e-commerce platforms to drive the digitalization process of the entire supply chain.

(3) Identifying issues in industry product e-commerce policies and standards: The research results effectively achieved the initial goal of uncovering issues within policies and standards. From the number of valid documents collected, it is evident that the strategic design and institutional framework for China's industrial product e-commerce are not yet well-established. The level of support from institutions and the digital ecosystem is relatively low, and there is insufficient institutional support for the digital transformation of small- and medium-sized enterprises (SMEs). Local governments have made relatively slow progress in implementing policies for industrial product e-commerce, with some regions even lagging behind. Although the national government has issued relevant policies, many regions have not yet formulated corresponding detailed rules or policy guidelines, leading to a lack of good cluster effects in the digital transformation of industrial SMEs and a weak e-commerce ecological environment. Moreover, despite significant advancements in the application of technology and platform construction in industrial product e-commerce, the literature reviewed shows a relative lack of discussion on "standards", "regulations", or "certification systems". This highlights the lag in standardization efforts within the field of industrial product e-commerce. The absence of unified standards has led to serious information asymmetry issues and increased transaction costs and, to some extent, has constrained the efficient operation and resource integration capabilities of the industrial product e-commerce market.

Based on our analysis and findings, government departments can further improve policies, while enterprises can further adjust and optimize their business strategies to jointly promote the healthy development of industrial product e-commerce.

As exploratory research, this study inevitably has certain limitations. First, owing to the relatively niche nature of industrial product e-commerce, the document sample size is relatively small, which may limit the generalizability and robustness of the findings to some extent. In addition, the sources used in this study are secondary data, including research reports and policy-type texts, so the conclusions obtained need further validation. More empirical studies are needed to verify and support them.

(1) To overcome these limitations, future studies may consider integrating more diverse data sources, such as news reports and transcripts of expert interviews, to enhance the broad applicability and reliability of the findings. Moreover, to further deepen understanding of industrial product e-commerce, specific case studies can be developed based on existing research. By selecting more representative industrial product e-commerce platforms or enterprises as research objects, combined with field research and in-depth interviews, we can dig deeper into the actual operation and closely combine theoretical analysis with practical operation. This will not only depict the structure and operation mechanism of industrial product e-commerce more specifically but will also provide more solid support for theoretical research.
(2) At the methodological level, future research can also explore the application of other advanced technologies and algorithms. In addition to existing text-mining techniques, machine learning and deep learning algorithms, such as BERT, GPT, and other natural language processing models, have demonstrated excellent text analysis capabilities when dealing with large-scale datasets, and some techniques can achieve fine-grained analysis even for small-scale datasets. Combined with word vector embedding techniques, these models can further improve the accuracy of understanding industry trends. In addition, the use of network analysis methods can more intuitively demonstrate the complex associations within and outside the industry, providing a new perspective for analyzing the industry chain synergistic effect of industrial product e-commerce.

In summary, through the extensive research above and the application of new technologies, future research can not only further enrich and improve the theoretical framework in the field of industrial product e-commerce but also provide more in-depth and practical decision-making support for the industry and policymakers and assist the continuous innovation and development of industrial product e-commerce.

Author Contributions: Conceptualization, Z.S. and Y.M.; methodology, Q.Z. and Y.M.; formal analysis, Q.Z.; writing—review and editing, G.W. and Z.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Basic Scientific Research Business Fees Project "Research on Machine-Readable Standards and Intelligent Application Technologies for the Clothing Industry" (No. 532023Y-10393) and Major Humanities and Social Sciences Research Projects in Zhejiang Higher Education Institutions (No. 2023QN077).

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.

References

1. Ben Youssef, A.; Dahmani, M. Examining the Drivers of E-Commerce Adoption by Moroccan Firms: A Multi-Model Analysis. *Information* **2023**, *14*, 378. [CrossRef]
2. Zhang, C.; Yang, Q.; Zhang, J.; Gou, L.; Fan, H. Topic Mining and Future Trend Exploration in Digital Economy Research. *Information* **2023**, *14*, 432. [CrossRef]
3. Martín-Gómez, A.M.; Agote-Garrido, A.; Lama-Ruiz, J.R. A Framework for Sustainable Manufacturing: Integrating Industry 4.0 Technologies with Industry 5.0 Values. *Sustainability* **2024**, *16*, 1364. [CrossRef]
4. Ocloo, C.E.; Xuhua, H.; Akaba, S.; Shi, J.; Worwui-Brown, D.K. The Determinant Factors of Business to Business (B2B) E-Commerce Adoption in Small-and Medium-Sized Manufacturing Enterprises. *J. Glob. Inf. Technol. Manag.* **2020**, *23*, 191–216. [CrossRef]
5. Zhang, Y.H.; Zhuang, Z.Z.; Li, Z.W. Can E-Commerce Promote Innovative Behavior in Traditional Manufacturing? *J. Quant. Tech. Econ.* **2018**, *35*, 100–115. (In Chinese)
6. Claycomb, C.; Iyer, K.; Germain, R. Predicting the Level of B2B E-Commerce in Industrial Organizations. *Ind. Mark. Manag.* **2005**, *34*, 221–234. [CrossRef]
7. Chen, M.L.; Chen, Y.F.; Lin, Q.Y. Research on Industrial Control Mode of Electronic Commerce under Industry 4.0 Background. *Manuf. Autom.* **2015**, *37*, 146–147+150. (In Chinese)
8. Waithaka, S.T.; Mnkandla, E. Challenges Facing the Use of Mobile Applications for E-Commerce in Kenya's Manufacturing Industry. *Electron. J. Inf. Syst. Dev. Ctries.* **2017**, *83*, 1–25. [CrossRef]
9. Tang, P.P.; Wu, L. A Research on the Effeet Mechanism of New Electronic Commerce to Theunderdeveloped Areas of China. *Chin. J. Manag.* **2014**, *11*, 1143–1149. (In Chinese)
10. Wang, D. Influences of Cloud Computing on E-Commerce Businesses and Industry. *J. Softw. Eng. Appl.* **2015**, *6*, 313–318. [CrossRef]
11. Al Mashalah, H.; Hassini, E.; Gunasekaran, A.; Bhatt, D. The impact of digital transformation on supply chains through e-commerce: Literature review and a conceptual framework. *Transp. Res. Part E* **2022**, *165*, 102837. [CrossRef]
12. Sharma, R.; Srivastva, S.; Fatima, S. E-Commerce and Digital Transformation: Trends, Challenges, and Implications. *Int. J. Multidiscip. Res. (IJFMR)* **2023**, *5*, 1–9.
13. Ren, L.; Wang, H.; Tang, Y.; Yang, C. AIGC for Industrial Time Series: From Deep Generative Models to Large Generative Models. *arXiv* **2024**, arXiv:2407.11480.
14. Juventia, S.D.; Jones, S.K.; Laporte, M.A.; Remans, R.; Villani, C.; Estrada-Carmona, N. Text Mining National Commitments Towards Agrobiodiversity Conservation and Use. *Sustainability* **2020**, *12*, 715. [CrossRef]
15. Puri, M.; Varde, A.S.; De Melo, G. Commonsense based text mining on urban policy. *Lang. Resour. Eval.* **2023**, *57*, 733–763. [CrossRef]
16. Tobback, E.; Naudts, H.; Daelemans, W.; De Fortuny, E.J.; Martens, D. Belgian Economic Policy Uncertainty Index: Improvement Through Text Mining. *Int. J. Forecast* **2018**, *34*, 355–365. [CrossRef]
17. Rao, G.K.; Dey, S. Decision Support for E-Governance: A Text Mining Approach. *Int. J. Manag. Inf. Technol.* **2011**, *3*, 73–91.
18. Talib, R.; Hanif, M.K.; Ayesha, S.; Fatima, F. Text Mining: Techniques, Applications and Issues. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*, 414–418. [CrossRef]
19. *GB/T 30698-2014*; E-Commerce Supplier Evaluation Criteria Quality Manufacturers. Standards Press of China: Beijing, China, 2014.
20. *GB/T 23031.2-2023*; Industrial Internet Platform Application Implementation Guide Part 2: Digital Management. Standards Press of China: Beijing, China, 2023.

21. *GB/T 42562-2023*; Industrial Internet Platform Selection Requirements. Standards Press of China: Beijing, China, 2023.
22. *GB/T 23031.5-2023*; Industrial Internet Platform Application Implementation Guide Part 5: Personalized Customization. Standards Press of China: Beijing, China, 2023.
23. *GB/T 23031.6-2023*; Industrial Internet Platform Application Implementation Guide Part 6: Service Extension. Standards Press of China: Beijing, China, 2023.
24. Ding, Y.; Teng, F.; Zhang, P.; Huo, X.; Sun, Q.; Qi, Y. Research on text information mining technology of substation inspection based on improved Jieba. In Proceedings of the 2021 International Conference on Wireless Communications and Smart Grid (ICWCSG), Aachen, Germany, 25–28 October 2021; pp. 561–564.
25. Qaiser, S.; Ali, R. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *Int. J. Comput. Appl.* **2018**, *181*, 25–29. [CrossRef]
26. Jin, X.; Zhang, S.; Liu, J. Word Semantic Similarity Calculation Based on Word2Vec. In Proceedings of the 2018 International Conference on Control, Automation and Information Sciences (ICCAIS), Hangzhou, China, 24–27 October 2018; pp. 12–16.
27. Krishna, K.; Murty, M.N. Genetic K-Means Algorithm. *IEEE Trans. Syst. Man Cybern. Part B* **1999**, *29*, 433–439. [CrossRef] [PubMed]
28. Sievert, C.; Shirley, K. LDAvis: A Method for Visualizing and Interpreting Topics. In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Baltimore, MD, USA, 27 June 2014; pp. 63–70.