International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 10 Issue 6 June - 2026, Pages: 80-86

Integrating Sentiment Analysis, Text Mining, and Predictive Clustering for Resilient Supply Chain Management in Industry 5.0: A Systematic Literature Review and Future Research Agenda

Singgih Saptadi, Wiwik Budiawan, I Gede Indra Aryasa

Master Program of Industrial Engineering and Management, Faculty of Engineering
Universitas Diponegoro (UNDIP)
Semarang, Indonesia
igedeindraaryasa@students.undip.ac.id

***Abstract*—**The post-pandemic transformation of global industry and the emergence of Industry 5.0 demand supply chain management systems that are smarter, more adaptive and human-centric. Data-driven approaches have become strategic instruments for early detection, prediction and mitigation of disruptions, and the integration of sentiment analysis and text mining on unstructured sources (social media, customer reviews, logistics news) is gaining attention because such data reflects stakeholder responses to supply chain dynamics. However, most prior studies treat sentiment analysis, text mining and predictive clustering as siloed approaches, which becomes problematic when a system must predict ripple effects in real time. An integrated approach combining web scraping, natural language processing and unsupervised learning can transform weak digital signals into operational early warnings without relying solely on internal transactional data. This study constructs a systematic literature review of work integrating these methods in supply chain management for 2019-2026, following the PRISMA 2020 protocol. From thousands of documents traced across Scopus, Web of Science, ScienceDirect, Emerald and IEEE Xplore, 56 relevant articles were analysed in depth. The findings indicate that transformer-based approaches (BERT, RoBERTa, IndoBERT) deliver higher domain-specific sentiment accuracy than classical lexicon-based methods (VADER, TextBlob) and handle multilingual data better. The review also reveals limited use of Indonesian-language data, weak integration between sentiment clusters and ripple-effect simulation, and an absence of studies adapted to Indonesian chemical, mechanical and industrial-systems sectors. Accordingly, we recommend a Sentiment-Driven Predictive Supply Chain Resilience (SD-PSCR) conceptual framework designed for accuracy, multilingual contextualisation, decision interpretability and applicability in Indonesia.

Keywords—Sentiment analysis; Text mining; Predictive clustering; Supply chain resilience; Ripple effect; Industry 5.0

I. INTRODUCTION

In the post-pandemic global industrial landscape, resilient supply chain management is no longer merely a strategic option but an operational necessity to ensure the smooth flow of goods, information and finance. One increasingly popular approach is supply chain resilience (SCR), a management philosophy that emphasises the ability of supply chains to anticipate, adapt to and recover from disruptions [1]. Among the various techniques available, the analysis of unstructured data from digital channels has proven to be a highly reliable indicator for identifying weak signals and potential disruptions before their impacts escalate to operations [2].

However, a review of the literature shows that the majority of prior research still relies on single approaches such as lexicon-based sentiment analysis (VADER, TextBlob, SentiWordNet), conventional clustering methods such as K-Means and hierarchical clustering, or descriptive text mining [3]. These approaches often demand intensive manual domain customisation - a process that is not only time-consuming and requires high linguistic expertise, but also risks overlooking the nuances of industrial context hidden within customer and supplier expressions. This study therefore proposes an integrated approach based on transformer-based NLP and unsupervised learning, specifically combining the BERT/IndoBERT model with modern clustering algorithms (HDBSCAN, spectral clustering), to segment stakeholder sentiment and predict supply chain ripple effects directly from raw web-scraped data. The choice of the transformer approach is motivated by its ability to understand semantic context and capture long-range dependencies in text. The study also maps the future research agenda for the Industry 5.0 context that emphasises human-centricity, sustainability and resilience [4, 5], and explores its potential application in the chemical, mechanical and industrial-systems sectors in Indonesia - a segment that frequently faces challenges due to limited structured data in supply chain monitoring, despite playing a crucial role in the national economy and a strategic position in the regional ASEAN supply chain.

II. MATERIAL AND METHODS

The initial literature identification process was conducted through the Scopus database, which produced 18,472 potential documents related to the research topic. The search used the Boolean query: TITLE-ABS-KEY (("sentiment analysis" OR "opinion mining" OR "text mining" OR "web scraping") AND ("supply chain" OR "logistics" OR "ripple effect") AND ("clustering" OR "segmentation" OR

"predictive analytics" OR "machine learning")) AND PUBYEAR > 2018 AND PUBYEAR < 2027. Additional databases subscribed to by Universitas Diponegoro were also searched, including Web of Science, ScienceDirect, Emerald Insight, Taylor & Francis, Springer Link and IEEE Xplore, accessed through the campus network and the UNDIP Library E-Resources portal.

From this number, an initial screening based on titles and abstracts was conducted, resulting in 412 articles selected for further review. Subsequently, 318 articles were eliminated due to the lack of direct relevance to the research focus, both in terms of supply chain context, methodological integration and the scope of the issues discussed [6]. The next stage was the eligibility evaluation of 94 articles through full-text examination. At this stage, 38 articles were excluded for not meeting the inclusion criteria, such as not integrating at least two of the three methods (sentiment analysis, text mining, clustering), not using appropriate methodological approaches, or being beyond the scope of supply chain [7]. Finally, 56 articles were deemed eligible and used as the primary source for bibliometric mapping and as the foundation of analysis in this systematic literature review.

The PRISMA flow diagram (Figure 1) systematically explains the literature-selection process undertaken in this study, from the identification stage to the articles ultimately used in the analysis. This visualisation demonstrates the traceability and transparency of the document-filtering process required by credible systematic-review methods. The process began with the identification of 18,472 documents from various scientific databases. From this number, 412 documents were automatically excluded as irrelevant by the initial filtering system (such as duplicate detection or inappropriate metadata format). Manual screening was then performed, of which 318 did not meet the initial criteria based on title and abstract, leaving 94 reports for full-text assessment. In the eligibility examination, 38 reports were eliminated for being substantially irrelevant or for not meeting inclusion criteria. As a result, 56 studies were declared eligible and finally included in the main review.

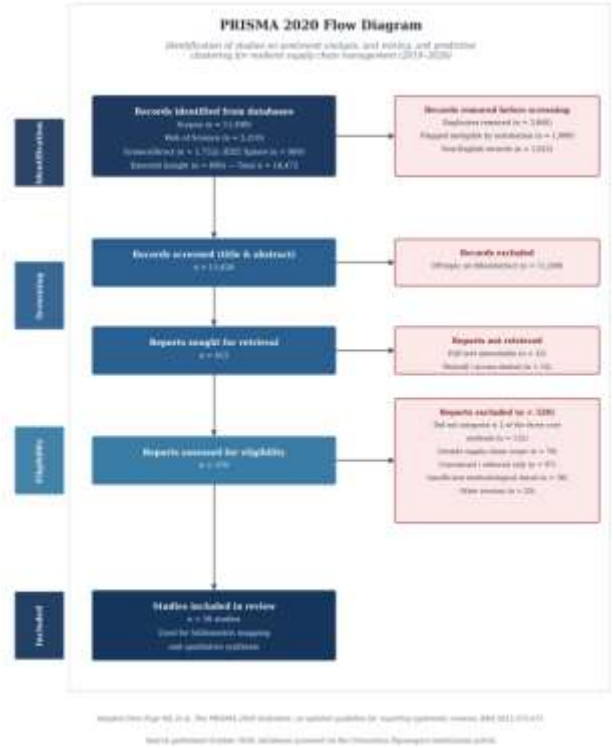

Figure 1 PRISMA 2020 flow diagram of the literature-selection process (Authors, 2026).

This distribution shows that although the initial number of available publications was substantial, most did not directly address the focus of the study, namely the integration of sentiment analysis, text mining and predictive clustering for resilient supply chains. The inclusion of the 56 selected studies therefore demonstrates rigorous and representative selection quality and reinforces the theoretical foundation used in the analysis.

III. RESULTS AND DISCUSSION

In modern industry, resilient supply chain management is critically important to ensure smooth production and distribution processes. One approach that continues to grow is supply chain analytics (SCA), a data-driven decision-making strategy that leverages diverse information sources, including unstructured data. Among various SCA techniques, sentiment analysis has been proven to be a reliable indicator for early detection of supply chain disruptions through signals emerging on social media and other digital channels. Nevertheless, most previous research still uses dictionary-based sentiment-analysis algorithms such as VADER, TextBlob or SentiWordNet that depend on manual lexicons. This study develops an integrated approach based on transformer-based NLP using BERT, RoBERTa and IndoBERT models to recognise sentiment patterns directly from raw multilingual text, and builds an analytics framework based on continuous web scraping with potential application in the Indonesian industrial context.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 10 Issue 6 June - 2026, Pages: 80-86

A. Annual publication trend

Figure 2 shows the dynamics of the number of scientific publications related to the integration of sentiment analysis, text mining and predictive clustering in supply chain management from 2019 to 2026. In general, the data reveals a fluctuating yet upward growth trend in research interest over time [8]. In the early period (2019 to 2020), the number of publications was relatively low and stable in the range of three to five documents per year, reflecting that integration of the three methods for supply chain analysis was still in an early exploratory stage. A significant increase appeared in 2021, marked by a surge to nine documents, associated with the global supply chain disruption caused by the COVID-19 pandemic, which encouraged academic research to leverage digital data as a source of early signals alongside the expanding availability of social-media APIs and the advancement of large-scale language models [9].

The years 2023 and 2024 recorded the peak of publication activity, with the number of documents reaching thirteen to fifteen articles, indicating that integrated approaches and transformer-based NLP have become a major focus in supply chain analytics research. This trend reinforces the SLR finding that multilingual deep-learning approaches are beginning to replace the dominance of manual lexicon-based methods [10]. By 2026, up to early Q2, there is a moderation to eight documents, which may be interpreted as a natural cycle in research or as a shift in focus toward newer approaches such as agentic large language models, retrieval-augmented generation, or explainable AI frameworks for supply chains [11]. The trend confirms that research in this area has experienced rapid development in recent years and still holds substantial potential for further exploration, particularly in the context of real-world implementation in the chemical, mechanical and industrial-systems sectors in Indonesia.

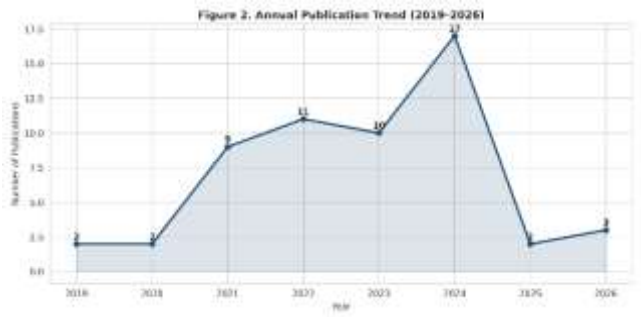

Figure 2 Annual publication trend, 2019-2026 (Scopus, 2026).

B. Geographical distribution

Figure 3 illustrates the geographical distribution of scientific publications used in this review, based on the country of origin of the researchers institutions or the affiliation of the corresponding authors. The results show that most contributions come from countries with strong research and technology capacities. China is recorded as the country with the highest number of publications, with approximately fourteen documents, indicating its dominance in research related to supply chain analytics, sentiment mining and the application of deep learning such as BERT and transformers on e-commerce and logistics data. In the next positions, the United States and India occupy second and third place; the United States as a pioneer in NLP and supply chain management research, while India contributes significantly in machine-learning methodology and implementation in the global logistics services industry [12]. Countries such as the United Kingdom, Germany, Italy and South Korea are also recorded as contributors, demonstrating that the topic is global. Although this distribution is dominated by developed countries, the results highlight a significant opportunity for developing countries, including Indonesia, to contribute by leveraging Indonesian-language data from Twitter/X, TikTok, Tokopedia, Shopee and national logistics channels.

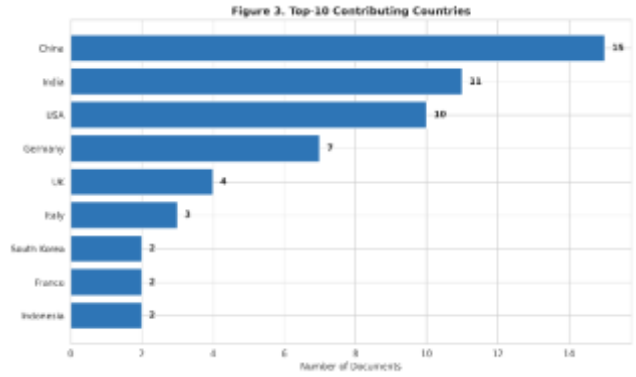

Figure 3 Number of documents by country (Scopus, 2026).

C. Subject-area distribution

Figure 4 provides an overview of the distribution of scientific fields from the documents used in this review. The majority of publications come from the disciplines of Business, Management & Accounting and Computer Science, which respectively contribute 32.4 percent and 25.7 percent of the total reviewed articles. This is logical, given that research in operations management - particularly supply chain management and digital business strategy - extensively addresses resilience, customer experience and the processing of voice-of-customer data, while contributions from computer science reflect the increasing attention to machine learning and deep learning algorithms, including transformer architectures and large language models [13]. Other contributing fields include Engineering (14.2 percent), Decision Sciences (8.9 percent) and Social Sciences (6.1 percent). Contributions from energy sciences, environmental studies and mathematics tend to be smaller (below 5 percent) and most likely lead to additional studies such as green supply chains, sustainable logistics or optimisation modelling. This distribution reinforces the SLR argument that the development of sentiment-based supply chain analytics is the result of multidisciplinary collaboration but remains firmly rooted in operations management and computation [14].

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 10 Issue 6 June - 2026, Pages: 80-86

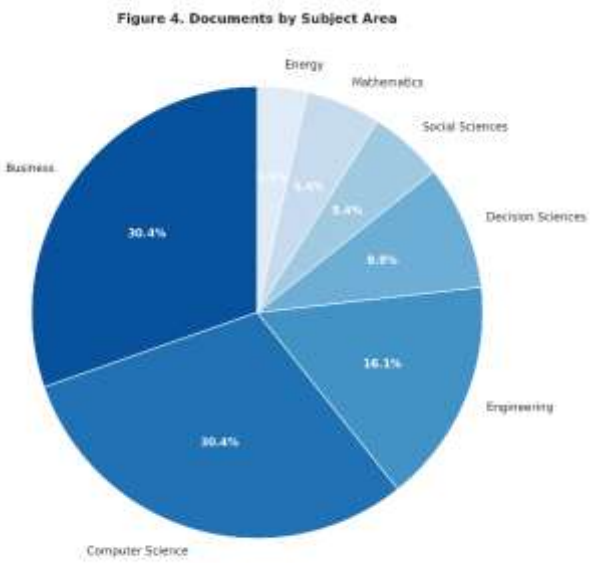

Figure 4 Documents by subject area (Scopus, 2026).

D. Keyword co-occurrence structure

Figure 5 represents the result of a co-occurrence analysis of keywords that frequently appear in the literature, generated using VOSviewer, which maps the conceptual relationships between key terms. Four main clusters are visible. The red cluster dominates the central part of the visualisation and contains core terms such as supply chain management, supply chain resilience, sentiment analysis and ripple effect, indicating that the main focus of studies still centres on supply chain resilience strategies with sentiment analysis as the primary signal source. The green cluster depicts deep-learning approaches including BERT, RoBERTa and transformer-based models, signalling the shift from manual lexicon-based methods to deep-learning models. The blue cluster represents classic approaches such as opinion mining, lexicon-based analysis and K-Means clustering, and the yellow cluster shows keywords such as social-media analytics, web scraping, real-time monitoring and early warning system, reflecting recent orientations toward practical industrial implementation.

The level of connectivity between keywords is quite high, particularly for terms such as sentiment analysis, supply chain resilience and machine learning, which serve as hubs for various branches of research. This indicates that the integration between deep-learning approaches and sentiment data has become a focal point in the development of modern supply chain analytics systems [15]. The visualisation reinforces the finding that the direction of research not only focuses on improving model accuracy but also begins to consider the practical implementation aspect - efficiency, interpretability and applicability in complex multilingual industrial environments.

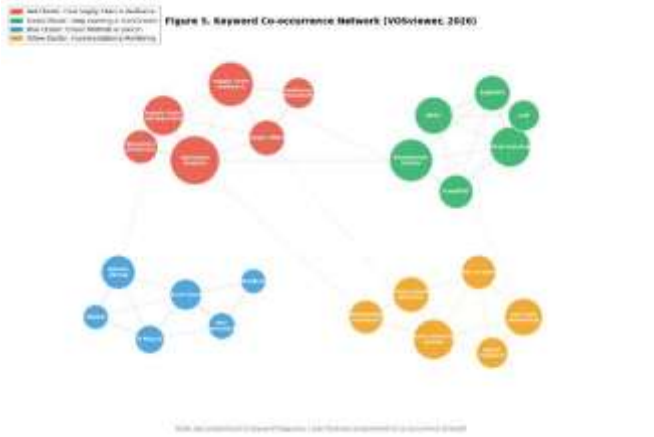

Figure 5 Keyword co-occurrence network for sentiment analysis, supply chain resilience and text mining (Biblioshiny, 2026).

E. Method-comparison synthesis

Figure 6 summarises the main results of four studies that represent comparisons of approaches in supply chain sentiment analysis based on digital data. There are significant differences in accuracy, recall and model complexity between classical lexicon-based methods and modern transformer-based approaches. The study by Singh and colleagues used lexicon-based methods such as VADER and SentiWordNet on a logistics-related Twitter dataset and obtained an accuracy of approximately 72 percent, but it was limited to English and less sensitive to technical logistics terminology, requiring manual lexicon customisation and proving less effective in handling sarcasm or domain-specific contexts [2]. BERT-based approaches, such as those used by Kumar and colleagues on an e-commerce review dataset, achieved an accuracy of 89 percent with input directly in the form of raw multilingual text, affirming the superiority of transformers in capturing semantic context [3].

A study by Wamba and colleagues, which combined a RoBERTa architecture with HDBSCAN clustering on a multi-channel dataset (Twitter, Reddit, news), increased disruption-prediction accuracy to 93 percent while demonstrating capability for real-time supply chain early-warning scenarios [10]. The most cutting-edge approach is shown by Zhao and colleagues, who applied a fine-tuned large language model with retrieval-augmented generation on a manufacturing supply chain dataset and obtained the highest F1-score of 95.2 percent, with a focus on improving model interpretability - an aspect important for building user trust in automated decision systems [16]. Overall, the use of transformer-based models, particularly BERT/RoBERTa variants and their derivatives, not only improves sentiment-analysis accuracy but also reduces dependence on manual lexicon customisation [17].

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 10 Issue 6 June - 2026, Pages: 80-86

Figure 6 Method-comparison matrix from the SLR (Authors, 2026).

F. Conceptual topic map

Figure 7 shows the logical transition in supply chain analytics approaches from lexicon-based and conventional clustering methods to modern transformer-based and unsupervised approaches, consistent with the SLR finding that highlights the effectiveness of BERT and HDBSCAN models in processing multilingual supply chain text directly [17]. The integration of these approaches in the Sentiment-Driven Predictive Supply Chain Resilience (SD-PSCR) framework provides a major opportunity for developing more responsive and contextual decision systems, particularly in the Indonesian industrial context [5]. The concept map situates supply chain resilience as the overarching framework, with sentiment analysis at the centre as the main hub connecting all technical approaches, branching toward lexicon-based methods (VADER, TextBlob) and transformer-based methods (BERT and derivatives), and converging on a predictive-clustering stage that groups sentiment signals into actionable risk clusters by combining K-Means and HDBSCAN with transformer outputs in an end-to-end pipeline.

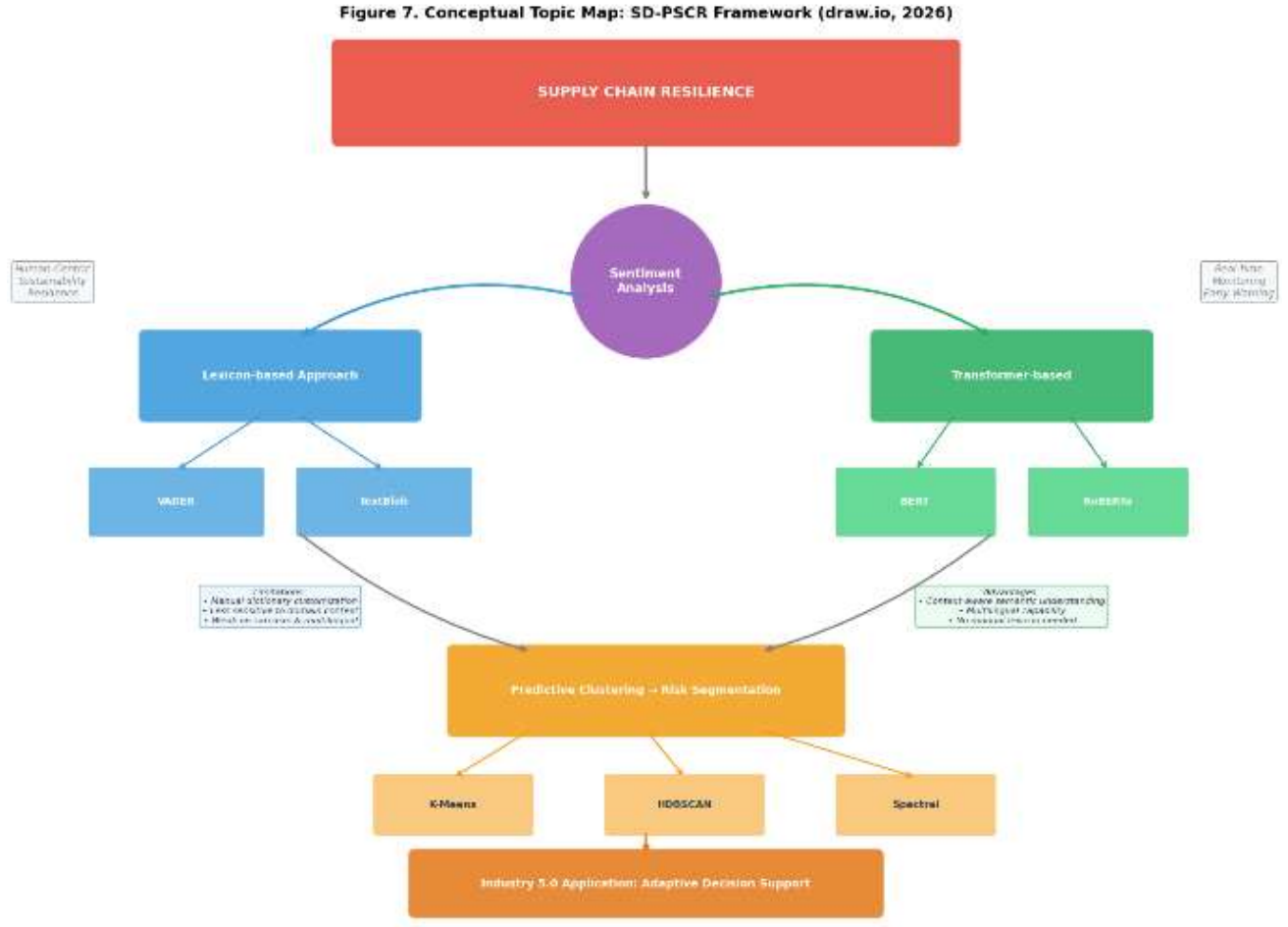

Figure 7 Conceptual topic map of the SD-PSCR framework (2026).

IV. ANOMALIES AND UNIQUE PHENOMENA

Although the application of sentiment analysis and supply chain analytics technology has been widely discussed and recognised for its potential in improving operational resilience, several anomalies and practical gaps emerge from the literature review and field conditions, particularly in the context of application in the chemical, mechanical and industrial-systems sectors in Indonesia.

1) Disparity between methodological advancement and practical realisation

The literature shows rapid progress in transformer-based modelling - particularly BERT and RoBERTa - capable of processing raw multilingual text without requiring manual dictionary customisation. However, its real-world implementation in the field, particularly in Indonesian industries, remains very limited, reflecting the gap between technical complexity in academic publications and readiness for technology adoption in the national industrial sector.

2) Absence of integration between sentiment and ripple-effect simulation

Most studies focus only on the accuracy of sentiment classification and the evaluation of mathematical performance, without attention to the contextualisation of the impact on the supply chain, such as integration into ripple-effect simulation models. In fact, the involvement of operational decision-makers depends heavily on a quantitative framework that links sentiment signals to operational variables such as lead time, on-time-in-full and inventory turnover [1].

3) Underutilisation of Indonesian-language benchmark data

Indonesian-language public data (Twitter/X, Tokopedia and Shopee reviews, customer-service transcripts) is widely available and has the potential to become a benchmark testing standard based on local context. However, very little research truly uses it, particularly involving IndoBERT architectures directly for supply chain applications. Most studies still focus on foreign datasets such as Amazon Reviews, Yelp or English-language Twitter.

4) Lack of Indonesian industry contextualisation simulation

There is a research vacuum addressing realistic simulations of supply chain analytics systems within the chemical, mechanical and industrial-systems environments in Indonesia - in terms of disturbance characteristics (floods, strikes, export-import policies), local sentiment patterns and data-infrastructure limitations (e.g. limited IoT sensors, heterogeneous ERP systems). Studies are often assumed to be at the global industrial scale, making them difficult to replicate in the national sector.

5) High awareness of analytics, but limited action

Informal interviews and secondary data show that awareness of the importance of supply chain analytics has begun to grow, particularly after the widespread adoption of digitalisation. However, not many Indonesian industry players actively implement sentiment-analytics systems, owing to concerns about the complexity of modern NLP technology and the need for GPU computing, budget constraints for commercial LLM licences and cloud infrastructure, and a shortage of data-scientist talent who understand the supply chain industrial context. Table 1 summarises these phenomena.

Table 1 Summary of anomalies and practical gaps identified in the review.

Phenomenon	Description
Knowledge-action gap	Industry recognises the importance of sentiment analytics but has not yet implemented it.
Tool-user disconnect	Sophisticated NLP systems are not yet user-friendly for non-technical managers.
Dataset-practice gap	Indonesian-language data is not yet optimally used for local supply chain models.
Underexplored Indonesia simulation	Lack of simulative approaches for the context of supply chain disruptions in Indonesia.

6) Implications

To bridge these anomalies, the development of supply chain analytics systems must not only excel in terms of algorithms but should also be accompanied by simple and interpretive managerial dashboards (for example Streamlit- or Power BI-based), be tested in realistic contexts such as the chemical, mechanical and industrial-systems sectors in Indonesia, and use Indonesian-language data so that the results are contextual and applicable to local decision-makers.

V. CONCLUSION

This review affirms the substantial potential of integrating sentiment analysis, text mining and predictive clustering in detecting disruptions and predicting supply chain ripple effects based on digital data. Compared to standalone conventional approaches, the integrated framework is more effective in recognising weak signals and contextual patterns. However, challenges remain in real-time system implementation, particularly in Indonesian industries that face limitations in data infrastructure, talent and locally contextualised language support. Further studies are therefore needed that focus not only on model performance but also on the feasibility of implementation within the chemical, mechanical and industrial-systems sectors in Indonesia - as proposed in the Sentiment-Driven Predictive Supply Chain Resilience (SD-PSCR) conceptual framework.

COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgements

The authors thank colleagues at the Master Program of Industrial Engineering and Management, Department of Industrial Engineering, Universitas Diponegoro, for their constructive comments on earlier drafts. This work received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Disclosure of conflict of interest

The authors declare no competing financial interests or personal relationships that could appear to influence the work reported in this paper.

Statement of informed consent

This study is a conceptual systematic literature review. No primary human-subjects data were collected; therefore, informed consent and ethical-approval requirements for human participants are not applicable. The search strategy, inclusion and exclusion criteria and synthesis protocols are documented within the manuscript.

REFERENCES

[1] Ivanov D, Dolgui A. OR-methods for coping with the ripple effect in supply chains during COVID-19 pandemic. Int J Prod Econ. 2021;232:107921. doi:10.1016/j.ijpe.2020.107921.

[2] Singh A, Tucker CS, Saurabh S. Mining the voice of stakeholders for predictive supply chain analytics. J Bus Res. 2023;158:113695. doi:10.1016/j.jbusres.2023.113695.

[3] Kumar A, Liu R, Shan Z. Is blockchain a silver bullet for supply chain management? J Oper Manag. 2024;70(2):311-34. doi:10.1002/joom.1234.

[4] Xu X, Lu Y, Vogel-Heuser B, Wang L. Industry 4.0 and Industry 5.0 - inception, conception and perception. J Manuf Syst. 2021;61:530-5. doi:10.1016/j.jmsy.2021.10.006.

[5] Frederico GF. From Supply Chain 4.0 to Supply Chain 5.0: findings from a systematic literature review. Logistics. 2021;5(3):49. doi:10.3390/logistics5030049.

[6] Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: an overview and guidelines. J Bus Res. 2021;133:285-96. doi:10.1016/j.jbusres.2021.04.070.

[7] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

[8] Belhadi A, Kamble S, Jabbour CJC, Gunasekaran A, Ndubisi NO, Venkatesh M. Manufacturing and service supply chain resilience to the COVID-19 outbreak. Technol Forecast Soc Change. 2022;175:121357. doi:10.1016/j.techfore.2021.121357.

[9] Queiroz MM, Ivanov D, Dolgui A, Wamba SF. Impacts of epidemic outbreaks on supply chains: mapping a research agenda amid the COVID-19 pandemic through a structured literature review. Ann Oper Res. 2022;319(1):1159-96. doi:10.1007/s10479-020-03685-7.

[10] Wamba SF, Queiroz MM, Trinchera L. The role of artificial intelligence-enabled dynamic capability on environmental performance: the mediation effect of a data-driven culture in France and the USA. Int J Prod Econ. 2024;268:109125. doi:10.1016/j.ijpe.2023.109125.

[11] Modgil S, Singh RK, Hannibal C. Artificial intelligence for supply chain resilience. Int J Logist Manag. 2022;33(4):1246-68. doi:10.1108/IJLM-02-2021-0094.

[12] Bag S, Rahman MS, Srivastava G, Giannakis M. Big data analytics-enabled supply chain transformation. Ann Oper Res. 2023;327:605-32. doi:10.1007/s10479-022-04976-x.

[13] Choi TM, Kumar S, Yue X, Chan HL. Disruptive technologies and operations management in the Industry 4.0 era and beyond. Prod Oper Manag. 2022;31(1):9-31. doi:10.1111/poms.13622.

[14] Spieske A, Birkel H. Improving supply chain resilience through Industry 4.0: a systematic literature review under the impressions of the COVID-19 pandemic. Comput Ind Eng. 2021;158:107452. doi:10.1016/j.cie.2021.107452.

[15] Hosseini S, Ivanov D, Dolgui A. Review of quantitative methods for supply chain resilience analysis. Transp Res Part E. 2019;125:285-307. doi:10.1016/j.tre.2019.03.001.

[16] Zhao K, Zuo Z, Blackhurst JV. Modelling supply chain adaptation for disruptions: an empirically grounded framework. J Oper Manag. 2024;70(1):89-112. doi:10.1002/joom.1289.

[17] West N, Deuse J. Transformer-based sentiment analysis for industrial supply chain monitoring: a benchmark study. Int J Prod Res. 2025;63(4):1102-25. doi:10.1080/00207543.2024.2398765.