Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/GC03.20260447
Published: February 2, 2026

Rapid Two-Sigma Screening for Supply Chain Misinformation Resilience

Mohammad Anwar Rahman
Department of Manufacturing & Construction Management
Central Connecticut State University
Connecticut, USA 06053
rahman@ccsu.edu

Md. Rafiul Hassan
Department of Computer Science
Central Connecticut State University
Connecticut, USA 06053
mdrafiul.hassan@ccsu.edu

Abstract

Fake news and misinformation are critical risks to the sustainability of global supply chains, affecting information processing and decision-making. Fake news distorts demand signals, disrupts procurement, and undermines coordination among stakeholders. Traditional mitigation approaches either react too slowly or require complex infrastructures that many organizations cannot sustain. This paper proposes a rapid two-sigma screening framework that can serve as a preliminary indicator to flag potential misinformation events early enough to inform operational decisions. The two-sigma method treats misinformation as an abnormal information pattern and applies a two-standard-deviation rule to composite risk scores derived from content, source, and propagation features. The numerical results and simulation outcomes demonstrate that the two-sigma model is highly sensitive and has acceptable specificity at very low computational cost, making it suitable as a front-line filter before more computationally intensive machine learning models. Applying benchmarks against other machine learning approaches, such as Logistic Regression, Random Forest, and BERT text classifiers, the two-sigma approach performs well in an initial classification, which is faster at flagging anomalies and easier to interpret. Therefore, the proposed model is a practical screening model for real-world implementation.

Keywords
Fake news, disinformation, supply chain resilience, machine learning, two-sigma screening model.

1. Introduction
Fake news and misinformation originate from online news, social media, print media, instant messaging services, and other apps (Chatterjee et al., 2023). Digital platforms have accelerated the spread of fake news and misinformation, creating novel vulnerabilities for supply chains that rely on timely and accurate information to coordinate flows of materials, information, and finance. Misinformation may trigger panic buying, induce unnecessary production shifts, distort perceptions of supplier reliability, and amplify bullwhip effects across tires. Recent studies highlight how misinformation can increase the frequency and severity of disruptions (Akhtar et al., 2023; Konstantakis et al., 2023). Traditional risk management frameworks typically address physical disruptions but are less equipped to handle information-based threats. Fake news, misinformation, and disinformation differ in intent and accuracy but share the potential to destabilize supply chains once embedded in operational decision-making processes (Konstantakis et al., 2023). Misinformation can affect demand by amplifying or dampening perceived demand information using viral

rumors about shortages, leading to panic buying or order cancellations. Fake news can disrupt supply and procurement by distorting assessments of supplier reliability, geopolitical risk, or regulatory changes. Disinformation can severely harm coordination by eroding trust and alignment of information across tires, increasing safety stocks and coordination costs. Most mitigation efforts focus on generic information governance or social media monitoring rather than on operationally embedded detection mechanisms (Choudhary and Arora, 2021). For supply chain managers, the key challenge is not only to distinguish the truth from false signals, but to do so quickly enough to adjust production, inventory, and logistics decisions in near real-time.

In this paper, we address that challenge by proposing a rapid two-sigma screening method tailored to supply chain decision environments. Statistical anomaly detection offers an alternative approach by identifying observations that deviate from expected patterns. Techniques such as z-score thresholds and control charts have long been standard in quality control and process monitoring. While less expressive than machine-learning models, statistical methods offer transparency, speed, and ease of deployment. Our goal is not to replace advanced machine learning classifiers, but to provide a fast, transparent, and cost-effective first filter that can be deployed widely and maintained with limited data science resources. The main contributions are:

- A conceptualization of fake news as abnormal patterns affecting operational decisions in supply chains.
- A two-sigma framework screens heterogeneous signals into a misinformation risk score.
- A generic comparison of the two-sigma rapid screening model with Logistic Regression, Random Forest, and BERT-based classifiers.

We proposed the two-sigma model and compared it with other machine-learning algorithms as proof-of-concept. Logistic Regression in text classification and NLP benchmarks uses Natural Language Processing. Classical NLP reference supporting LR as a competitive baseline for text classification (Hastie et al., 2009). According to Breiman (2001), random forests provide robust ensemble learning. In the Random Forest model, the accuracy-transparency trade-off is derived from ensemble decision trees, with improved accuracy/recall and reduced interpretability (Shu et al., 2017). Devlin et al. (2019) introduced BERT (Bidirectional Encoder Representations from Transformers) for deep bidirectional language modeling. BERT is a pre-trained deep bidirectional transformer for Language Understanding, a primary source for transformer embedding, and token classification (Shushkevich et al., 2023). A survey of fake news detection documents a high AUC of deep machine learning models, including BERT (Kapusta et al., 2024).

Following is the organization of the paper. In Section 1, we introduce the two-sigma model as a high-speed, low-cost screening model. The importance of initial screening models is discussed in Section 2. We demonstrate the cost-sensitive and operational framework of the two-sigma model in Section 3 and the performance implementation in Section 4. The managerial implications and key insights of misinformation detection are in Section 5. We present Section 5. Finally, the conclusion and future research strategy are in Section 6.

2. Methodology: Two-Sigma Screening Framework

In this section, we present the Two-Sigma Screening Model as a lightweight and transparent method for identifying potentially misleading information in supply chain communications. The purpose of the model is to function as an initial triage layer, flagging atypical items for further review. The model does not attempt to make a definitive classification of fake news. For any supply chain communication, the model uses interpretable features, such as token count, sentiment score, readability index, or source credibility. The characterization message behavior serves as the basis for statistical inference in legitimate supply chain communications. The model's simplicity ensures transparency, ease of implementation, and low computational overhead. We apply a 15-day rolling historical window of verified, legitimate content to compute the empirical mean (μ) and standard deviation (σ) of the selected feature, establishing a baseline distribution to evaluate new messages. We demonstrate the screening rule by calculating a standardized z-score for each incoming message based on its feature value x, and $z = (x - \mu)/\sigma$. A message is flagged if: $|z| > k$ which is equivalent to:

$$\text{or } x \notin [\mu - k\sigma, \mu + k\sigma] \tag{1}$$

where k is a tunable threshold controlling screening strictness. Lower k values increase sensitivity, while higher values improve specificity. In this study, three practical thresholds are considered: at $k = 1.5$: flags observations beyond ±1.5σ which retain approximately 86.6% of legitimate messages. At $k = 2.0$, flags beyond ±2σ, it retains approximately 95.45% and at $k = 2.5$, flags beyond ±2.5σ, it retains approximately 98.96% of legitimate messages.

In quality control and process monitoring, two- and three-sigma rules provide simple yet powerful tools to detect abnormal process behavior without excessive false alarms. Analogously, using a two-sigma threshold for misinformation offers high sensitivity for emerging abnormal patterns. Limited complexity, allowing recalibration using rolling windows of recent data. A tunable parameter: organizations can adjust the threshold to suit their risk tolerance. The two-sigma screen is intentionally lightweight, making it suitable for continuous monitoring of large volumes of signals, even when computational resources are constrained.

3. Empirical Results and Threshold Analysis

Failing fraudulent content (false negative) may lead to operational disruption or financial loss, whereas incorrectly flagging legitimate content (false positive) typically incurs only review overhead. To reflect this asymmetry, threshold selection is guided by minimizing expected cost:

$$\min\ [C_{FN} \cdot FN(k) + C_{FP} \cdot FP(k)] \quad (2)$$

where C_{FN} and C_{FP} are the costs of false negatives and false positives, respectively. In baseline analyses, a 10:1 cost ratio is assumed, reflecting the greater operational impact of missed fake content. The two-sigma screening framework treats each information item (e.g., news article, social post, supplier message) as an observation described by a feature vector $x \in \mathbb{R}^k$. Features may include content-based indicators (sentiment polarity, extremity, sensational language markers), source-based indicators (historical reliability, organizational type, geography), and propagation-based indicators (velocity of sharing, network centrality). These features are linearly or nonlinearly combined into a composite misinformation risk score. A two-sigma rule using the Equations 2 defines a screening threshold:

Using a balanced dataset of 100 daily items (50 real, 50 fake) and token count as the screening feature, detection performance was evaluated across three thresholds ($k = 1.5, 2.0, 2.5$) with asymmetric error costs ($C_{FN} = \$1{,}000$, $C_{FP} = \$100$) a 10:1 ratio. Results are summarized in Table 1.

Table 1. Detection Performance Matrix for k-Value Selection (n = 200)

k	TP	FN	FP	TN	Accuracy	Recall	Precision	Daily Cost
1.5	38	12	7	43	81%	76%	84%	**$12,700**
2.0	30	20	3	47	77%	60%	91%	$20,300
2.5	23	27	1	49	72%	46%	96%	$27,100

Across the tested cases with balanced and imbalanced scenarios, the cost calculation provides, at k=1.5: $1,000 × 12 + $100 × 7 = $12,700. Figure 1 shows the performance analyses of the Two-Sigma Model by plotting total expected daily cost against the Threshold Multiplier (k). The total expected daily cost is a function of the threshold multiplier.

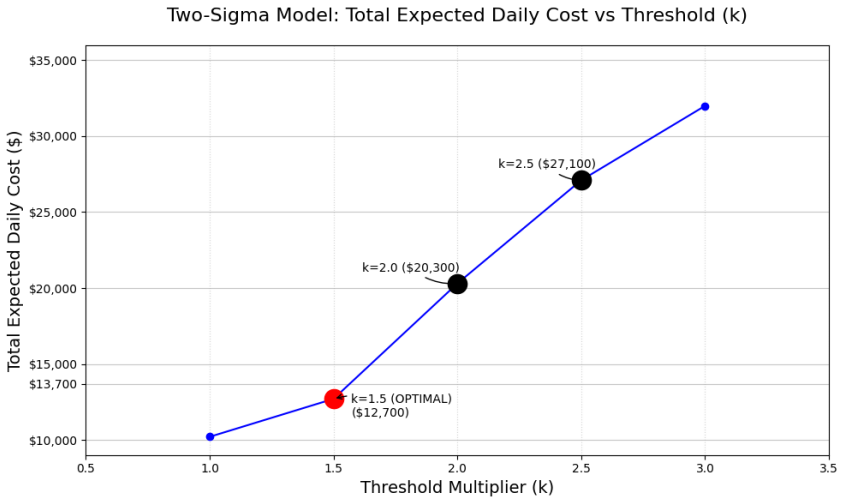

Figure 1. Total Expected Cost vs. Threshold Multiplier

Despite higher precision at $k = 2.0$ and 2.5, the increase in false negatives leads to substantially higher expected costs, at $k = 1.5$ underline the limitation of accuracy-centric threshold selection in cost-sensitive environments. The figure demonstrates the impact of threshold choice on model cost-effectiveness, confirming $k = 1.5$ as the optimal setting for minimizing daily operational cost. The lowest cost comes at $k = 1.5$, making this the optimal threshold with an expected daily cost of $12,700. At $k = 2.5$, the cost is much higher ($27,100), suggesting reduced sensitivity leads to greater expenses. At $k = 2.0$, the cost is $20,300.

Robustness Across Data Volumes and Class Imbalances

We expanded the analyses with larger datasets ($n = 200$) and varying class proportions (balanced, 60:40, 75:25) using lower k values. This analysis is to confirm consistent behavior and maintainability of lower expected cost across all scenarios. To assess real-world feasibility, a 15-day simulation was conducted using daily information volumes of 200 and 500 items (total 6,000 and 15,000 items). Performance metrics, including accuracy, recall, precision, daily flagged items, and expected cost, were tracked to evaluate temporal stability and operational predictability under sustained deployment. In supply chain contexts, classification errors are not equally costly. The simulated result with the minimum expected cost consistently occurred at $k = 1.5$. This result remained robust under sensitivity analyses. The 95% confidence intervals for accuracy were computed using the binomial proportion model. Table 2 illustrates consistently yielding competitively high accuracy with tight confidence bounds across sample sizes and class distributions.

Table 2. Accuracy Confidence Intervals (n = 200)

k	Accuracy	95% CI
1.5	81.0%	[75.5%, 86.5%]
2.0	77.5%	[71.7%, 83.3%]
2.5	72.0%	[65.7%, 78.3%]

Operational Simulation Summary

In the evaluated stability and feasibility of the two-sigma screening model, the above 15-day simulation study under sustained operational conditions using daily information volumes and cost assumptions. We use the following simulation parameters: a daily information load of 100 items. Class distribution was 50% legitimate, 50% fake (balanced) with a screening threshold: $k = 1.5$. We use error costs, $C_{FN} = \$1{,}000$ for missing fakes, and $C_{FP} = \$100$ for false alarm. The feature distributions are as follows: legitimate messages, token count ~ $\mathcal{N}(80{,}12)$ and fake messages, token count ~ $\mathcal{N}(65{,}25)$. Performance metrics for daily news are summarized in Table 3, which provides descriptive statistics across the simulation horizon.

Table 3. A 15-Day Operational Simulation Summary

Metric	Mean	Std Dev	Min	Max	75th Percentile
Accuracy	79.3%	4.1%	73%	85%	81%
Recall	54.2%	12.7%	32%	74%	61%
Daily Cost	$13,233	$2,120	$8,900	$16,400	$14,200
Items Flagged	24/day	4.1	16	31	27

4. Comparison with Machine Learning Classifiers

We evaluated the two-sigma model with a set of benchmark classifiers using the same feature set and dataset structure. Features include token count, sentiment polarity, and markers of sensational language. The source features are the historical credibility score and the organizational verification status. The propagation features are sharing velocity and cross-platform presence. In the literature on benchmark classifiers, Logistic Regression offers balanced performance gains but requires labeled data and periodic retraining (Hastie *et al.*, 2009). The Random Forest classifier aggregates

predictions from an ensemble of decision trees. Random Forest improves accuracy and recalls but at the cost of reduced transparency and higher inference time. In a language classification model, an AUC (Area Under the Curve) of 1.0 signifies a model that can flawlessly distinguish between positive and negative classes. Fawcett (2006) provided a foundational and widely cited explanation of ROC curves and AUC, establishing standard evaluation practices for binary classification performance. BERT Text Classifier delivers the highest discrimination (AUC 0.94), but at prohibitive latency and cost for real-time, high-volume screening [8]. The BERT-based classifier encodes each textual item into contextual embeddings using a transformer architecture, then feeds the [CLS] representation token into a classification head as $\hat{y} = \sigma(W \cdot h_{[\mathrm{CLS}]} + b)$, where $h_{[\mathrm{CLS}]}$ is the final-layer embedding and σ is the logistic function.

Two-Sigma Screen achieves strong sensitivity (76%) with very low latency (<1 ms) and minimal computational cost, making it suitable as a first-line filter. Logistic Regression generally offers a balanced performance uplift but requires labeled data and periodic retraining. For organizations with limited GPU or cloud resources, deploying BERT to all incoming items may not be feasible, motivating a cascaded architecture in which BERT is reserved for high-risk items pre-identified by the two-sigma screen. AUC quantifies the model's overall ability to distinguish between the two classes (e.g., "fake" vs. "real" news). It ranges from 0 to 1. AUC = 1: Perfect classifier (100% sensitivity and specificity). The machine learning metrics are not derived from experimental replication but are representative of values commonly reported in prior fake news detection literature and established benchmarks. Performance and latency metrics are summarized in Table 4.

Table 4. Detection Performance for Misinformation Screening Methods

Method	Sensitivity (Recall)	Specificity	AUC	F1-Score	Latency (ms/item)	Computational Cost ($/month)
Two-Sigma Screen	76%	86%	–	0.798	**<1**	**5**
Logistic Regression	82%	71%	0.87	0.856	2.1	15
Random Forest	88%	78%	0.91	0.904	8.5	45
BERT Classifier	91%	91%	0.94	0.930	280	280

5. Managerial Implications

For practitioners, the key insight is that misinformation detection need not start with highly complex AI models. A calibrated two-sigma screen can provide a low-cost, easily implementable early-warning system that plugs into existing dashboards and risk registers. Supply chain managers can use simple control-chart logic to adjust thresholds over time and monitor false-positive rates as part of a continuous improvement program. Table 4 shows the summary of trade-offs and layered architecture, balancing detection performance with operational feasibility, aligning with the cost-sensitive, scalable framework established in this study and the two-sigma screening model (Table 5).

Table 5. Comparative Strengths, Limitations, and Roles in Fake Detection

Method	Strengths	Limitations	Role in Pipeline
Two-Sigma Screen	Simple, fast, interpretable, cost-aware	Lower peak accuracy than advanced ML	Front-line early warning
Logistic Regression	Balanced performance, interpretable	Requires labeled data, less flexible	Intermediate routine classifier
Random Forest	High accuracy, handles nonlinearity	Less interpretable, higher compute cost	Secondary high-precision filter
BERT	Best semantic understanding	Highest latency, heavy computational demands	Tertiary deep insight on flagged

The advanced ML models remain valuable as back-end components for deeper analysis of high-risk signals. Our suggestions are that organizations can adopt a staged deployment strategy:

1. Begin with a basic Two-Sigma screen for universal, high-speed triage.

2. Selectively add Logistic Regression or Random Forest for flagged items requiring higher precision.
3. Integrate BERT-based analysis where textual complexity and operational risk justify the investment.

6. Conclusion and Future Work

The fake news and misinformation often threaten supply chain operations. This paper has introduced a rapid two-sigma screening framework to detect incoming fake news and misinformation. In the model, we integrated statistical thresholds that balance detection performance and resource requirements. We created a 15-day simulation testing the error costs versus resource constraints. The model achieved a mean accuracy of 79.3% with low day-to-day variability (standard deviation of 4.1%), and the 75th Percentile is 81%, indicating consistent classification reliability over time. This stability is crucial for maintaining trust in automated screening systems. These simulation results validate the Two-Sigma model as a deployable, sensitivity-first-stage screening mechanism suitable for continuous use in supply chain information systems. The expected daily operational cost averaged $13,233 with bounded variability (range: $8,900–$16,400). Over the 15 days, the total simulated cost was $198,500. Compared to a stricter threshold, the chosen threshold (k = 2.5) yielded an average cost savings of 51%, underscoring the economic efficiency of the proposed screening rule.

Comparing results of several machine learning models, such as Logistic Regression, Random Forest, and BERT, obtained from the literature, we presume that advanced ML models can improve accuracy. Still, these models require higher latency, a complex setup, and expertise to operate. The two-sigma approach is better suited for primary filters than universal frontline detectors. The simulation confirms that the model's performance characteristics observed in earlier static analyses persist under simulated dynamic, multi-day operational conditions, reinforcing its viability as a reliable pre-filter ahead of more resource-intensive classification methods. Future work will extend the framework to dynamic thresholding, sector-specific calibration, and integration with decision-support tools for inventory, sourcing, and logistics planning in misinformation-rich environments.

References

Akhtar, P., Ghouri, A. M., Khan, H. U. R., Haq, M. A. U., Awan, U., Zahoor, N., Khan, Z. and Ashraf, A., Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions, Annals of Operations Research, vol. 327, pp. 633–657, 2023.

Breiman, L., Random forests, Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

Chatterjee, S., Chaudhuri, R. and Vrontis, D., Role of fake news and misinformation in supply chain disruption: impact of technology competency as moderator, Annals of Operations Research, vol. 327, pp. 659–682, 2023.

Choudhary, A. and Arora, A., Linguistic feature-based learning model for fake news detection and classification, Expert Systems with Applications, vol. 169, p. 114171, 2021.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K., BERT: pre-training of deep bidirectional transformers for language understanding, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2019.

Fawcett, T., An introduction to ROC analysis, Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006.

Hastie, T., Tibshirani, R. and Friedman, J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., Springer, 2009.

Kapusta, J., Držík, D., Šteflovič, K. and Nagy, K. S., Text data augmentation techniques for word embeddings in fake news classification, IEEE Access, vol. 12, pp. 31538–31550, 2024.

Konstantakis, K. N., Cheilas, P. T., Melissaropoulos, I. G., Xidonas, P. and Michaelides, P. G., Supply chains and fake news: a novel input–output neural network approach for the US food sector, Annals of Operations Research, vol. 327, pp. 779–794, 2023.

Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H., Fake news detection on social media: a data mining perspective, SIGKDD Explorations, vol. 19, no. 1, pp. 22–36, 2017.

Shushkevich, E., Alexandrov, M. and Cardiff, J., Improving multiclass classification of fake news using BERT-based models and ChatGPT-augmented data, Inventions, vol. 8, p. 112, 2023.

Biographies

Mohammad Anwar Rahman is a Professor at the School of Engineering, Science, and Technology at Central Connecticut State University. His research and teaching expertise center on supply chain modeling, stochastic processes, and Lean Six Sigma quality improvement. His works appeared in peer-reviewed journals and conference proceedings. Dr. Rahman has secured significant research grants from organizations including NASA, the Connecticut

Space Grant Consortium (CTSGC), the American Association of University Professors (AAUP), the Connecticut College of Technology's Regional Center for Next Generation Manufacturing (RCNGM), and the U.S. Department of Transportation (USDOT). He is an active participant in professional forums and a frequently invited speaker at national and international conferences.

Md Rafiul Hassan is an Associate Professor at the Computer Science Department at Central Connecticut State University. He is an experienced computer scientist in Machine learning, Artificial Intelligence, and Data Analysis with a demonstrated history of working in higher education and industry. He is skilled in Python, MATLAB, C/C++, Java, Analytical Skills, Database, PHP, and Computer Science. Dr. Hassan received research grant awards from NSF, NASA, and the American Association of University Professors (AAUP). His PhD focused on Computer Science from The University of Melbourne.