Journal of
Theoretical and Applied
Electronic Commerce Research

MDPI

Article

Enhancing Supply Chain Resilience in Textile SMEs: A Human-Centric Customer-to-Manufacturer Framework Using Public E-Commerce Data

Chien-Chih Wang *, Yu-Teng Hsu and Hsuan-Yu Kuo

Department of Industrial Engineering and Management, Ming Chi University of Technology, New Taipei City 243303, Taiwan
* Correspondence: ieccwang@mail.mcut.edu.tw; Tel.: +886-2-2908-9889

Abstract

Upstream textile small and medium-sized enterprises (SMEs) frequently exhibit constrained supply chain resilience owing to persistent information latency and structural dependence on downstream orders. To address these challenges, this study develops and validates a customer-to-manufacturer (C2M) intelligence framework that enables data-driven production planning using publicly available e-commerce data. The framework incorporates ethically compliant acquisition of consumer demand signals, semantic translation of unstructured market data into textile engineering attributes, machine-learning-based demand forecasting, and human-centric decision support. Utilizing 3.87 million consumer comments from 127,846 product listings, a Neural Boosted Tree model with entity embeddings for textile attributes was constructed. This model achieved a mean R^2 of 0.921 in cross-validation, surpassing benchmark methods. Consumer comment volume was validated as a proxy for sales activity, facilitating demand estimation. Forecasts were translated into production guidance using Monte Carlo simulation and a decision dashboard. In a 12-month field study at a Taiwanese dyeing SME, implementation resulted in a 28% reduction in inventory value, a 31% decrease in dye lot changeovers, and a 16% increase in capacity utilization. This research extends the C2M paradigm from downstream retail contexts to upstream textile SMEs, proposes an integrated and operationally feasible intelligence framework for resource-constrained manufacturers, and demonstrates how digital intelligence can enhance supply chain resilience while supporting, rather than replacing, human decision-making. The results indicate that upstream textile SMEs can leverage publicly visible e-commerce signals to enhance production planning responsiveness, minimize inventory exposure and dye-lot disruptions, and strengthen resilience to demand uncertainty through planner-centered digital decision support.

Keywords: customer-to-manufacturer; supply chain resilience; Industry 5.0; human-centric AI; public data stream analytics; textile SMEs

Check for updates

Received: 17 February 2026
Revised: 13 April 2026
Accepted: 15 April 2026
Published: 17 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The global apparel industry is undergoing rapid transformation, driven by athleisure consumption, shorter product cycles, and rising demand for personalization. These dynamics intensify demand volatility and compress planning horizons, particularly for functional garments, where performance attributes and aesthetics require frequent updates [1]. Consequently, forecast-driven and order-lagged planning can become structurally misaligned

with rapidly changing consumer preferences, increasing the risk of overproduction and obsolete inventory. Building on supply chain research emphasizing the alignment between product characteristics and supply chain design, agile and responsive planning approaches are widely recognized as critical in fashion-oriented markets.

As manufacturing paradigms evolve from Industry 4.0 to Industry 5.0 [2], with increased focus on human-centricity, sustainability, and resilience, upstream textile operations such as dyeing and finishing continue to encounter significant constraints related to information latency and dependence on downstream orders [3]. This structural lag intensifies demand distortions as signals propagate upstream, a phenomenon known as the bullwhip effect, in which minor fluctuations at the consumer level are progressively magnified throughout the supply chain, ultimately compromising operational stability [4]. Taiwan's upstream textile sector offers a particularly instructive context for examining this issue. Although the industry is recognized for its strength in functional fabric innovation and its critical role in global performance textile supply networks, most upstream firms remain resource-constrained and operate primarily on a make-to-order basis. In contrast to large enterprises that invest in proprietary data ecosystems, upstream small and medium-sized enterprises (SMEs) lack systematic external sensing capabilities to detect early demand shifts, resulting in reactive capacity planning and increased vulnerability to post-pandemic volatility [5].

Taiwan represents a particularly suitable empirical context for this study for four primary reasons. First, Taiwan remains a globally significant supplier of functional and sustainable textiles, with exports reaching US$6.1 billion in 2024, demonstrating ongoing international relevance and substantial exposure to evolving end-market demand. Second, the sector is characterized by a dense network of specialized upstream and midstream manufacturers, many of which face resource constraints and have limited visibility into downstream market signals. Third, major e-commerce platforms in Taiwan generate extensive, publicly accessible consumer engagement data in apparel-related categories, creating a robust environment for evaluating whether external market signals can inform upstream production intelligence. Fourth, post-pandemic uncertainty in export markets and order visibility positions Taiwan's textile industry as a directly relevant context for investigating resilience-oriented planning under demand volatility. Collectively, these factors establish Taiwan's upstream textile sector as an analytically rigorous and practically significant setting for examining how customer-to-manufacturer (C2M) intelligence can enhance SME supply chain resilience.

Therefore, the challenge lies not only in technological limitations but also in organizational factors: effectively converting fragmented market signals into actionable production intelligence without displacing experienced planners. The customer-to-manufacturer (C2M) paradigm presents a promising approach by linking consumer behavior to upstream manufacturing decisions [6]. E-commerce environments generate high-frequency signals, including product attributes, consumer feedback, and engagement traces, which can serve as leading indicators of demand sensing [7]. Nevertheless, a persistent industrial integration gap remains in terms of transforming raw, noisy, and unstructured public signals into granular, attribute-level intelligence (e.g., color, material, and function) suitable for upstream planning [8]. Furthermore, the literature frequently underemphasizes the human-centric requirements for SMEs; practical digital transformation necessitates interpretable decision support that reduces information latency and bridges the semantic gap between consumer language and engineering parameters.

This study addresses these challenges by developing and field-validating an upstream attribute-level consumer-to-manufacturer (C2M) decision-support framework using publicly available Shopee e-commerce data in Taiwan. Specifically, we pursued three objectives:

(1) to develop a robust and platform-friendly data acquisition pipeline for standardizing real-time consumer signals; (2) to benchmark machine-learning models for high-cardinality textile attributes and develop a Neural Boosted Tree approach with entity embeddings to improve generalization; and (3) to operationalize probabilistic forecasts into a human-centric "Traffic-Light" dashboard to reduce cognitive load in shop-floor decision-making. The framework was deployed in a 12-month longitudinal field study at a Taiwanese dyeing SME, where its implementation was associated with audited, observed improvements in firm records, including reduced inventory value, fewer dye-lot changeovers, and higher capacity utilization. Collectively, this study demonstrates a replicable pathway for upstream SMEs to enhance resilience through low-cost digital intelligence, aligning operational decision-making with Industry 5.0 principles.

This study is structured around three primary research questions. First, it investigates whether publicly available e-commerce consumer signals can be ethically acquired and transformed into attribute-level demand intelligence suitable for upstream textile production planning. Second, it examines whether a neural-boosted tree architecture with entity embeddings outperforms conventional tree-based forecasting models when applied to high-cardinality, noisy textile demand data. Third, it explores whether a probabilistic, human-centric decision support system can operationalize such forecasts to deliver measurable improvements in supply chain resilience for small and medium-sized enterprises (SMEs) in real-world longitudinal deployments. The research contributions are threefold. The study extends the customer-to-manufacturer (C2M) paradigm from finished-product retail forecasting to upstream, component-level manufacturing, specifically dyeing and finishing operations, thereby providing an end-to-end framework applicable to resource-constrained SMEs. It introduces a domain-adapted neural-boosted tree model with entity embeddings, validated against five benchmark architectures using 3.87 million consumer records across 127,846 product listings under high-cardinality textile data conditions. Finally, it designs and deploys a Bayesian Monte Carlo probabilistic framework, operationalized as a traffic-light dashboard, that translates forecast uncertainty into interpretable shopfloor signals and embodies Industry 5.0 principles of human-centricity and resilience.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature on supply chain resilience, C2M, and human-centric decision support. Section 3 details the system architecture and ethical data acquisition methodology. Section 4 presents benchmarking results and insights from industrial deployment. Section 5 discusses theoretical and practical implications. Section 6 concludes with limitations and future research.

2. Literature Review

2.1. The Digital Divide and Resilience in Upstream SMEs: An Industry 5.0 Perspective

The textile and apparel sector is characterized by extreme demand volatility, rapid product proliferation, and short product life cycles [9]. Classical time-series techniques perform adequately under stable conditions but fail to capture the nonlinearity in modern fast-fashion demand [10]. A critical structural disadvantage for upstream manufacturers is the "industrial information gap." Unlike large brands with integrated enterprise resource planning (ERP) ecosystems, upstream small and medium-sized enterprises (SMEs) often operate in a "digital shadow, " relying on downstream order lags as their primary demand signal [4,11]. Application-oriented studies have reported similar challenges in SME settings [12]. Recent studies have shown that SMEs face specific dual hurdles of technological gaps and organizational immaturity [13]. Unlike large enterprises, SMEs cannot afford complex commercial solutions to ensure data security. As Yilmaz et al. argue, these firms require "low-cost, accessible digital solution areas" (e.g., the Shoestring approach) rather than high-end infrastructure to initiate digitalization [14]. This lack of digital capability

directly impacts supply chain resilience. Wang et al. identify that information barriers are among the most significant obstacles to adopting resilience practices [15]. Without affordable tools to anticipate "unforeseen disruptions," SMEs remain reactive [16]. In the context of Industry 5.0, the objective shifts from pure automation to human-centric systems that protect both workers and businesses [17]. However, current forecasting remains anchored in internal history and fails to leverage external intelligence to enhance resilience.

2.2. External Sensing: From Unstructured Streams to Actionable Intelligence

The current state of customer-to-manufacturer (C2M) research is structured around three dimensions: accomplished milestones, unresolved challenges, and driving forces necessitating further progress. In terms of achievements, the rapid expansion of e-commerce has produced substantial digital footprints, such as sales volumes, reviews, and search queries, which function as open-source intelligence (OSINT) for manufacturing applications [18]. The C2M paradigm utilizes this external sensing capability to facilitate demand-driven manufacturing processes [6,19,20]. Additionally, application-oriented studies have demonstrated approaches for real-time integration [21]. However, the operationalization of C2M at the upstream component manufacturing level remains unresolved. Existing research primarily addresses forecasting at the finished-product and retailer levels, and lacks platform-compliant data acquisition pipelines, semantic translation mechanisms for engineering taxonomies, and probabilistic uncertainty frameworks suitable for small- and medium-sized enterprise (SME) shop-floor planners. The impetus for resolving these issues is driven by three factors. First, Industry 5.0 mandates enhancing digital resilience across SME supply chain tiers. Second, the increasing availability of high-density public e-commerce repositories has enabled cost-effective external sensing. Finally, post-pandemic demand shocks have exposed the inadequacy of reactive order-lagged planning for upstream manufacturers.

However, operationalizing this data for upstream manufacturing presents significant engineering challenges. Accessing social media or e-commerce data alone is insufficient. Tortorella et al. caution that while social media enhances information sharing, it can also lead to "information overload" and noise if not properly filtered, potentially hindering problem-solving routines [22]. A critical "semantic granularity gap" remains; raw e-commerce data are unstructured and noisy, lacking the standardized engineering attributes (e.g., color codes and material composition) required by shop-floor systems. Few studies have addressed the challenge of semantic interoperability–disaggregating noisy web text into attribute-level intelligence actionable by component manufacturers [18]. Bridging this semantic gap is essential for converting noisy external streams into the low-cost resilience tools envisioned by Yilmaz et al. [14].

2.3. Machine Learning Architectures for High-Cardinality Industrial Data

Machine learning offers superior capabilities for modeling complex nonlinear relationships compared to traditional statistical methods [23]. Within the domain of structured tabular data, tree-based ensembles, such as XGBoost, have established dominance owing to their robustness [24]. However, in the textile context, machine learning is typically applied to clean internal data. Applying these models to crawled, high-dimensional data introduces the challenge of high cardinality—for example, thousands of distinct color/material combinations. Standard encoding methods often yield sparse matrices and suboptimal performance. Prior studies have highlighted the sensitivity of ensemble methods to data sparsity and noise in industrial forecasting contexts [25]. Recently, neural-boosted architectures and entity embeddings have emerged to address this challenge, offering superior representation learning for high-dimensional categorical data [26]. Furthermore, effec-

tive industrial decision-making necessitates not only point forecasts but also the quantification of uncertainty. Integrating Bayesian approaches to account for the stochastic nature of consumer-derived sales proxies remains relatively underexplored in customer-to-manufacturer (C2M) research, particularly in upstream manufacturing contexts, where direct demand observations are unavailable.

2.4. Synthesis and Research Gap

The literature indicates a clear shift toward data-driven manufacturing; however, a gap persists between advanced Customer-to-Manufacturer (C2M) concepts and the operational realities of small and medium-sized enterprises (SMEs). As summarized in Table 1, existing studies provide limited treatment of the challenges associated with ethical data acquisition and semantic translation in resource-constrained environments. Consequently, three interrelated limitations have been identified:

Table 1. Key limitations in existing research and how this study addresses them.

Limitation Type	Description	Representative Studies	How This Study Addresses It
Information asymmetry & external sensing gap	Reliance on closed, lagged internal information loops, such as ERP orders, due to the absence of external demand signals, limits resilience against disruptions.	[4,11,15]	Develops a platform-friendly pipeline for extracting and standardizing publicly available consumer demand signals as real-time market sensors to inform upstream production planning.
Semantic interoperability & knowledge translation	Forecasts remain at the finished-product level, failing to provide attribute-level specifications necessary for dye-lot planning, owing to semantic gaps between consumer data and manufacturing parameters.	[19,22,27]	Proposes a standardized Color–Material–Function taxonomy to translate unstructured consumer information into attribute-level probabilistic forecasts for human-centric decision support.
Robustness under high-cardinality data regimes	Limited benchmarking of forecasting models was conducted under conditions of noisy, high-cardinality categorical features, and validation was lacking in real SME manufacturing environments.	[24–26]	Evaluated neural-boosted tree models incorporating entity embeddings to address data sparsity, and assessed robustness through a longitudinal industrial deployment.

This study develops and validates an end-to-end framework for transforming real-time consumer signals into actionable production intelligence, offering a cost-effective pathway for small and medium-sized enterprises to enhance Industry 5.0-oriented resilience.

2.5. The Role of Industry 5.0 in Textile Supply Chains

Industry 5.0, as defined by the European Commission, represents a paradigm shift beyond the efficiency-focused automation of Industry 4.0 by emphasizing human well-being, ecological sustainability, and operational resilience in manufacturing [3]. Although Industry 4.0 introduced transformative digital infrastructure such as cyber-physical systems, digital twins, and Internet of Things (IoT) connectivity, its implementation in small and medium-sized enterprise (SME)-dominated sectors like textiles has often revealed a risk of human displacement, as automation advances more rapidly than workforce adaptation [5]. In

contrast, Industry 5.0 positions artificial intelligence to augment human capabilities rather than replace them, aligning with human-in-the-loop manufacturing philosophies. Within the textile and apparel sector, initial Industry 5.0 applications have focused on worker safety and ergonomic support. For example, Pistolesi et al. demonstrated a smartwatch-LiDAR integrated system to assess musculoskeletal disorder risks among textile workers, directly applying the human-centricity pillar of Industry 5.0 on the shop floor [17]. Recent studies indicate that human–cobot collaboration can improve production flexibility while preserving meaningful roles for skilled operators in assembly-intensive manufacturing contexts [28,29]. Regarding sustainability, Kim et al. employed digital twin technology to develop an intelligent analysis module that enables real-time, continuous monitoring and analysis of dyeing parameters, thereby optimizing textile dyeing processes [30].

At the supply chain level, Industry 5.0 is transforming both the conceptualization and implementation of resilience in textile and apparel networks. Wieland and Durach contend that achieving supply chain resilience in the fashion industry necessitates a shift from reactive "bounce-back" strategies to proactive adaptation and reconfiguration capabilities. This transition aligns with the human-centric sensing architectures proposed in Industry 5.0 [31]. Choi et al. similarly demonstrated that integrating real-time consumer sentiment data into upstream supply chain decisions reduces demand amplification effects and enhances coordination across tiers in fast-fashion supply networks. This evidence suggests that external digital signals can strengthen resilience at the supply chain level, rather than only at the machine or process level [32]. From the perspective of SMEs, Bag et al. found that Industry 5.0-aligned digital practices, such as human-in-the-loop analytics and low-cost sensing, can substantially improve supply chain visibility and adaptive capacity in resource-constrained manufacturing firms, without requiring the large-scale infrastructure investments typically associated with Industry 4.0 [33]. Collectively, these findings suggest that the textile sector's transition to Industry 5.0 should move beyond internal process optimization to include the entire upstream sensing and decision-making pipeline, especially for SMEs facing persistent information asymmetry.

A major limitation of current Industry 5.0 textile research is its predominant focus on internal process optimization or final-stage assembly, leaving the upstream dyeing-and-finishing segment, where demand-signal latency is most pronounced, largely unaddressed. Additionally, the integration of external consumer-generated market signals with Industry 5.0-aligned decision frameworks has not yet been demonstrated for textile SMEs. This study addresses these gaps by proposing a C2M intelligence framework that operationalizes all three Industry 5.0 pillars: human-centricity, through a traffic light dashboard supporting planner judgment; sustainability, via forecast-guided batch consolidation to reduce dye lot changeovers, water, and chemical use; and resilience, through proactive demand sensing associated with an estimated six to eight week reduction in information latency between market signals and upstream production actions.

3. Methodology

This study introduces an end-to-end C2M intelligence framework designed to bridge the semantic gap between unstructured public e-commerce signals and shop-floor production parameters. Addressing the specific resilience needs of SMEs, the system is designed as a low-cost, modular solution that converts noisy external data into attribute-level production intelligence. The system architecture, illustrated in Figure 1, comprises four integrated modules: (1) publicly visible data acquisition (a platform-friendly pipeline); (2) semantic interoperability and ontology mapping, standardizing consumer language into engineering taxonomies; (3) high-cardinality attribute learning using neural boosted trees with entity embeddings; and (4) probabilistic decision support using Bayesian inference that enables

risk-aware production planning. Bayesian inference is a statistical method that systematically updates uncertainty estimates as new data become available, producing probability distributions that quantify forecast risk instead of relying on single-value predictions.

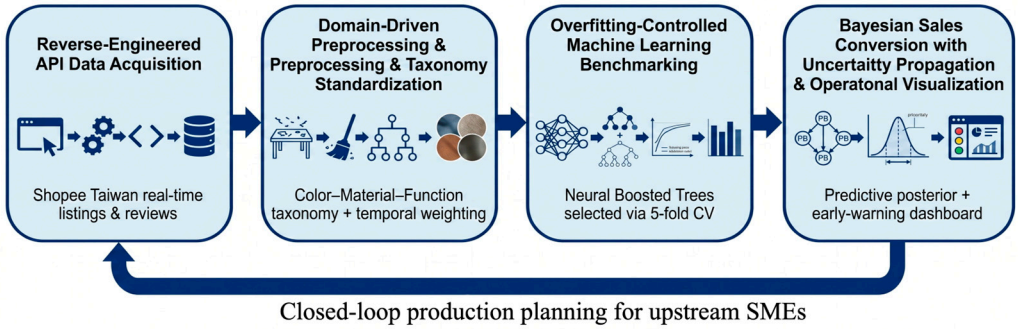

Figure 1. Architecture of the Human-Centric C2M Intelligence Framework for resilient SME production planning.

3.1. Ethical Acquisition of Publicly Visible E-Commerce Data

To support demand forecasting and production planning for functional textile products, we developed a modular data acquisition pipeline to collect publicly available product and review information from Shopee Taiwan. The pipeline employs a server-friendly, ethics-aware design that prioritizes lightweight data exchange, conservative request scheduling, and avoids full-page rendering. This approach reduces network overhead while maintaining data completeness and robustness to changes in the front-end interface. The three-stage architecture (Figure 2) integrates browser-based enumeration of public product listings, structured ingestion of lightweight item metadata, and retrieval of publicly visible consumer reviews.

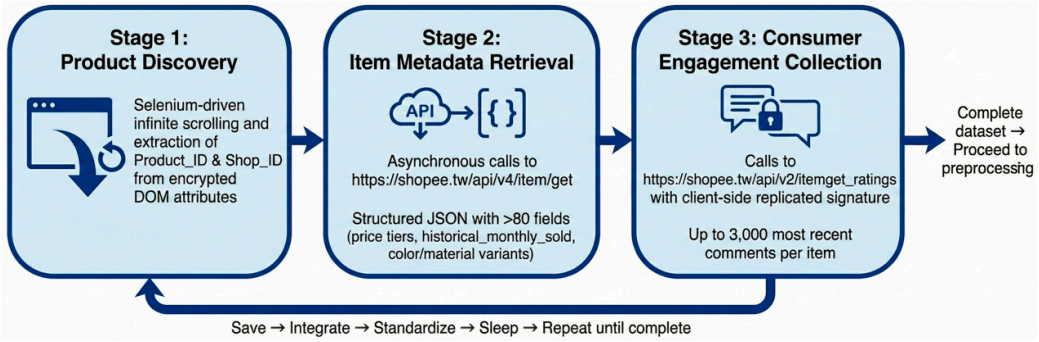

Figure 2. Three-stage pipeline for the ethical acquisition of publicly visible e-commerce data.

- Stage 1: Product discovery through the enumeration of public listings via a browser-based system.

Products were enumerated across four functional categories—Moisture-wicking, Cooling, Thermal, and Windproof/Waterproof—reflecting common functional classifications in the apparel and textile markets. Due to the dynamic loading of listings and potential exceedance of standard pagination limits, browser automation was employed to replicate typical user interactions, such as incremental scrolling, and to record network responses generated by the platform's public listing interface. From these publicly visible responses, product and shop identifiers were extracted to construct a comprehensive sampling frame that remained resilient to minor front-end updates.

- Stage 2: Structured metadata ingestion from public interface items.

The collected identifiers were used to retrieve structured item-level metadata via publicly accessible item-detail interfaces, yielding compact, JSON-formatted records (Figure 3). This metadata includes variant-level attributes (e.g., color and size) and time-indexed historical sales indicators, supporting semantic mapping and demand forecasting for functional textile products. Compared with browser-only retrieval, this approach substantially reduces computational and bandwidth overheads. All requests were scheduled using conservative rate-limiting and back-off strategies to ensure platform stability.

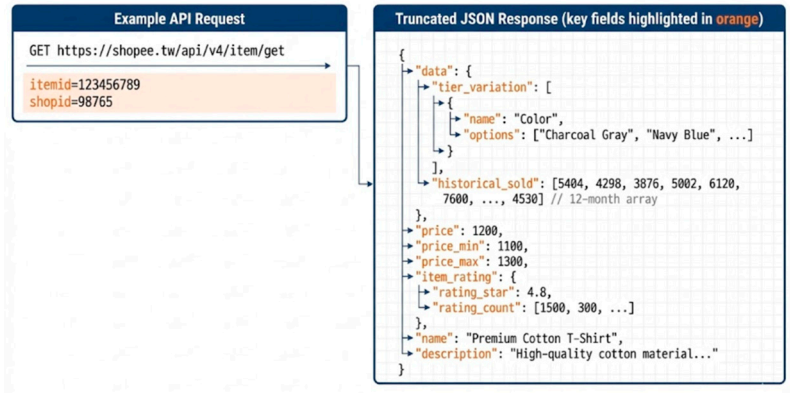

Figure 3. Structure of publicly accessible item-level metadata used in this study.

- Stage 3: Retrieval of publicly visible consumer review signals.

To capture downstream consumer engagement as a proxy for market demand, publicly visible product reviews were collected via the same interfaces accessible to standard web clients. Data collection employed session-aware scheduling and request throttling to prevent disruptions to platform services. Only information necessary for analysis–specifically, review text and aggregated rating indicators was retained, and no personally identifiable information was stored. For reproducibility, we report the data fields used, sampling criteria, and cleaning procedures, omitting implementation details that could be misused to circumvent platform protections.

In this study, we collected only publicly available information and did not access private user accounts or restricted content. Our data collection complied with platform access constraints and used conservative rate limiting to minimize platform load.

We did not conduct interviews, surveys, or controlled experiments with individual participants in this phase. It should be noted that the domain-expert review described in Section 3.2 (Stage 2) constitutes an internal professional validation procedure conducted by research team members, rather than human-subjects data collection, and is therefore governed by distinct ethical considerations. Consequently, no personally identifiable information was collected or processed, and institutional review board (IRB) approval was not required for the procedures described herein. Owing to platform terms, we cannot publicly release raw records. However, we provide derived features and processing logic to support reproducibility.

3.2. Domain-Driven Semantic Ontology Mapping and Data Harmonization

Given the high heterogeneity of raw e-commerce text, we engineered a rigorous four-stage pipeline (Figure 4) to transform unstructured descriptions into a standardized Industry 5.0 Semantic Ontology, creating a modeling-ready dataset of 51,072 observations.

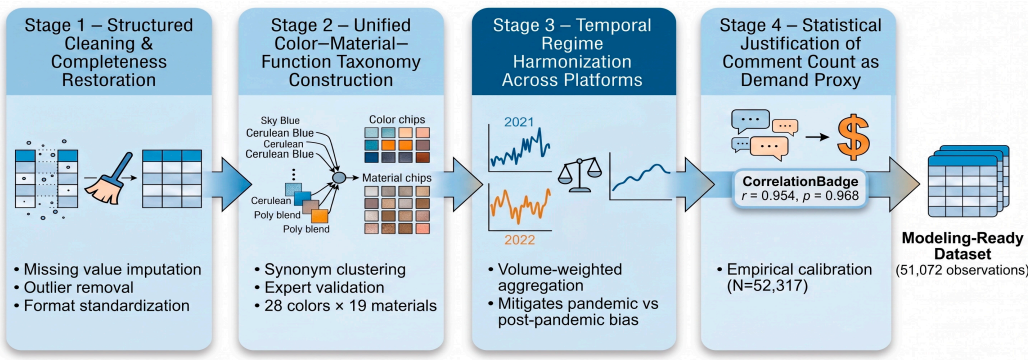

Figure 4. A four-stage semantic ontology mapping pipeline converts raw e-commerce data into a standardized industrial dataset.

- Stage 1: Algorithmic Completeness Restoration (Knowledge-Based Imputation)

To address data sparsity (18.4% attribute-level missingness), a hybrid recovery strategy was employed. A curated domain-specific lexicon, comprising 480 functional textile keywords, was combined with fuzzy string-matching algorithms (Levenshtein ratio ≥ 0.88) to recover missing attribute values from product titles. Listings with more than 50% missing critical attributes were algorithmically excluded, balancing data quality and sample coverage and resulting in a controlled attrition rate of 11.9%.

- Stage 2: Semantic Ontology Construction

To address extreme lexical variation, such as 47 distinct strings representing "Nylon," a unified color–material–function ontology was developed. FastText embeddings were employed for semantic clustering, followed by an internal domain-knowledge review. Two research team members, each with over 10 years of textile industry experience, independently assessed the canonical color–material–function triplets for semantic accuracy and manufacturing relevance. This review was an internal professional validation. No external participants were involved, and no additional IRB oversight was required (see Section 3.1). This process reduced the attribute cardinality from 12,847 raw strings to a canonical set of 2128 unique triplets (28 colors, 19 materials, and four functions). Dimensionality reduction retained 99.3% of the variance in principal component analysis (PCA) projections, minimizing information loss and ensuring semantic interoperability.

- Stage 3: Volume-Weighted Temporal Harmonization

To mitigate the impact of non-stationary market regimes induced by external shocks (e.g., pandemic-related demand distortions observed between 2021 and 2022), direct temporal concatenation of annual data was deliberately avoided. Instead, a volume-weighted temporal aggregation was employed to harmonize item-level demand signals across years.

$$y_i^{harmonized} = \frac{w_{2021} y_{i,2021} + w_{2022} y_{i,2022}}{w_{2021} + w_{2022}} \quad (1)$$

where $y_{i,t}$ denotes the observed demand proxy for item i in year t, and w_t represents the total platform sales volume in year t. By weighting annual observations according to overall market activity, this aggregation reduces the influence of transient demand anomalies and preserves the relative demand structure across items. As a result, the training signal remains robust to macro-environmental shifts, thereby enhancing the stability and resilience of the downstream demand modeling process.

- Stage 4: Ground-Truth Proxy Validation via Transient Data Visibility

A critical engineering challenge was validating "Comment Count" as a reliable proxy for "True Sales." We used a calibration subset of 52,317 items observed during a transient

system update window, during which exact sales figures were publicly available in an unredacted format. The empirical validation results (Table 2 and Figure 5) demonstrate a near-perfect log-log correlation (Pearson $r = 0.954$, Spearman $\rho = 0.968$).

Table 2. Empirical validation statistics for comment count as a sales demand proxy (calibration sample, N = 52,317).

Statistic	Value	Interpretation
Pearson correlation (log–log)	0.954	Extremely strong linear relationship
Spearman rank correlation	0.968	Outstanding monotonicity, robust to heavy tails
Mean comment-to-sale ratio ($\hat{\lambda}$)	0.0437	≈22.88 comments
95% prediction interval for λ	[0.0412, 0.0465]	High cross-sectional and temporal stability

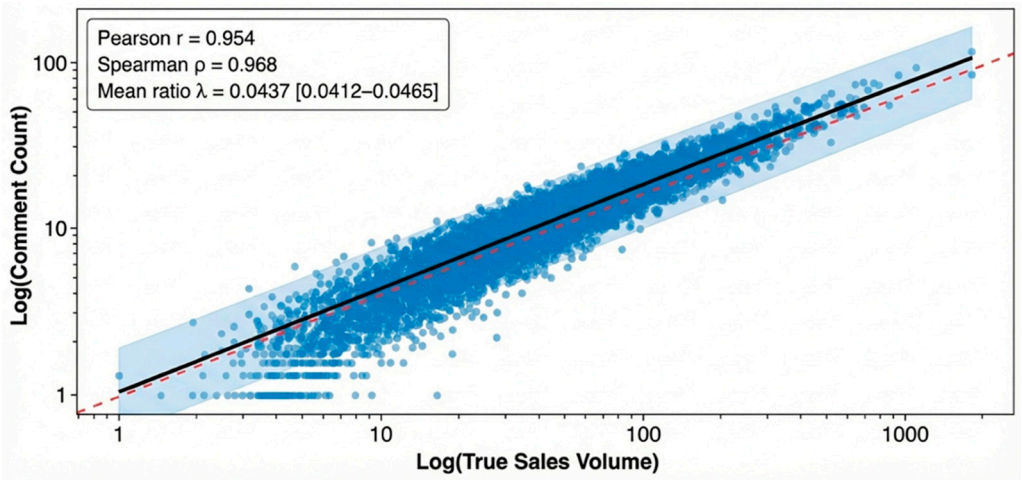

Figure 5. A log–log scatter plot of true sales volume versus comment count (N = 52,317) is presented, incorporating a LOESS regression fit and a 95% prediction interval.

As shown in Figure 5, the narrow prediction band confirmed the high fidelity of the proxy. The derived conversion ratio (mean $\hat{\lambda} = 0.0437$) exhibited high cross-sectional stability (95% CI: [0.0412, 0.0465]). Based on this compelling evidence, we formally adopted the cumulative comment count per standardized attribute triplet as the target variable y_i. This approach provides a scientifically validated alternative to the coarse rounded sales figures typically available on public platforms.

3.3. High-Cardinality Feature Learning and Model Benchmarking

We formulated the attribute-level demand-forecasting task as a supervised regression problem. Let $D = \{(x_i, y_i)\}_{i=1}^{N}$ be the dataset where y_i is the temporally harmonized comment count for a unique product entity. The high-dimensional feature vector x_i comprises continuous variables (price, product age, sentiment score, and competitor density) and high-cardinality categorical attributes (e.g., specific color/material codes). A critical challenge in textile informatics is the "curse of dimensionality" introduced by categorical features. Standard encoding methods (e.g., One-Hot Encoding) often yield sparse matrices that degrade tree-based learners. To address this, we implemented a systematic benchmarking protocol that compared five architectures spanning the spectrum of ensemble learning. Modeling was conducted independently for each of the four functional product categories to account for the market heterogeneity.

We evaluated representatives from three paradigms using identical stratified train–validation splits (80%/20%) to isolate algorithmic efficacy.

(1) Baseline: The CART model served as an interpretability baseline.

(2) Bagging: A bootstrap forest was selected to evaluate the variance-reduction capabilities of bagging on noisy e-commerce data.
(3) Boosting: We compared gradient-boosted trees with XGBoost. Although XGBoost is the industry standard, its default handling of high-cardinality data through sparse splits can be susceptible to overfitting in "small-N, large-P" regimes.
(4) Hybrid Architecture: We introduced Neural Boosted Trees (via H2O Driverless AI). This architecture hypothesizes that integrating Entity Embeddings (to capture semantic relationships in categorical data) with the residual learning of gradient boosting can improve the capture of interaction effects in sparse, high-cardinality datasets.

In the initial screening phase (Table 3), a critical "validation collapse" was observed in the pure boosting methods. Although XGBoost achieved a near-perfect training fit ($R^2 > 0.93$), its validation performance degraded sharply (average $\Delta R^2 = 0.58$). This pattern aligns with prior findings that tree-boosting methods may overfit under sparse, high-cardinality categorical regimes [34].

Table 3. Initial screening performance of the five benchmarked algorithms (default hyperparameters).

Model	Metric	Moisture-Wicking	Cooling	Thermal	Windproof/ Waterproof
XGBoost	R^2	0.966 (0.673)	0.960 (0.151)	0.931 (0.264)	0.945 (0.322)
	RMSE	10.064 (28.382)	12.740 (49.717)	26.973 (81.689)	4.721 (19.920)
Gradient Boosted Tree	R^2	0.898 (0.768)	0.915 (0.725)	0.638 (0.328)	0.818 (0.377)
	RMSE	17.937 (23.940)	18.588 (28.277)	61.486 (78.035)	8.563 (19.088)
Neural Boosted	R^2	0.883 (0.863)	0.862 (0.890)	0.761 (0.800)	0.839 (0.655)
	RMSE	18.593 (18.357)	23.773 (17.926)	50.016 (42.572)	8.047 (14.198)
Bootstrap Forest	R^2	0.871 (0.781)	0.839 (0.671)	0.841 (0.851)	0.801 (0.707)
	RMSE	19.527 (23.251)	25.668 (30.953)	40.786 (36.701)	8.949 (13.100)
CART	R^2	0.763 (0.529)	0.485 (0.352)	0.561 (0.647)	0.755 (0.494)
	RMSE	26.440 (34.072)	45.887 (43.420)	67.827 (56.530)	9.937 (17.199)

In contrast, Neural Boosted Trees and Bootstrap Forest demonstrated superior architectural robustness, maintaining training–validation gaps below 0.12. The incorporation of entity embeddings within the neural architecture contributed to regularization by leveraging the inherent sparsity of textile attribute taxonomies. Following this screening, the two models were selected for intensive hyperparameter optimization using a 50-trial Tree-structured Parzen Estimator (TPE) search. Comparative results indicated that Neural Boosted Trees consistently outperformed Bootstrap Forest across cross-validation folds, with statistically significant differences (paired t-test, $p < 0.01$). The final model performance is reported in the Results section.

3.4. Probabilistic Decision Support for Supply Chain Resilience

3.4.1. Bayesian Uncertainty Quantification and Sales Conversion

While the harmonized comment count serves as a robust demand proxy, effective resilience planning requires forecasts expressed in actual sales units, such as meters of fabric. Relying on publicly displayed "historical monthly sales" data is insufficient because these figures are dynamically rounded and lack temporal persistence. To address this limitation, we developed a statistically rigorous conversion mechanism calibrated against the "Ground Truth" subset observed during the transient data visibility window (detailed in Section 3.2).

We model the relationship between observed comments N_c and true sales N_s as a stochastic process. Assuming a constant conversion rate λ within the category, the likelihood is modeled as a Binomial distribution

$$N_c \sim \text{Binomial}(N_s, \lambda) \tag{2}$$

To minimize subjective bias, we employ Jeffrey's non-informative prior for the conversion rate:

$$\lambda \sim \text{Beta}(0.5, 0.5) \tag{3}$$

By conjugacy, the closed-form posterior distribution for λ is derived as:

$$\lambda | N_c, N_s \sim \text{Beta}(0.5 + N_c, 0.5 + N_s - N_c) \tag{4}$$

Calibration on the ground-truth dataset yielded a posterior mean of $\hat{\lambda} = 0.0437$ (approximately 1 comment per 22.88 sales), with a tight 95% Highest Density Interval (HDI) of [0.0412, 0.0465].

To operationalize this for forecasting, we reject simple point estimates in favor of Full Uncertainty Propagation. For a given Neural Boosted Trees prediction $\hat{y}_{i,m+1}$ (predicted comments), the corresponding sales forecast $\hat{S}_{i,m+1}$ is generated as a probability distribution via Monte Carlo sampling (K = 1000 draws):

$$\hat{S}^{(k)}_{i,m+1} = \frac{\hat{y}_{i,m+1}}{\lambda^{(k)}}, \text{ where } \lambda^{(k)} \sim \text{Beta}_{posterior} \tag{5}$$

This probabilistic formulation transforms a single scalar prediction into an empirical distribution of possible sales outcomes. This represents a critical resilience mechanism, as it enables the downstream decision support system to compute risk metrics—such as "Stockout Probability" or "conservative 95% Lower Bound"—rather than relying on brittle deterministic averages.

3.4.2. Human-Centric Operationalization: The Traffic-Light Early-Warning Interface

To bridge the gap between probabilistic forecasting outputs and shop-floor execution, the Monte Carlo ensembles generated in Section 3.4.1 are integrated into an interactive, human-centric decision support system (DSS) (Figure 6). The core of this interface is a rule-based traffic-light protocol designed to reduce cognitive load by translating distributional uncertainty into interpretable operational signals.

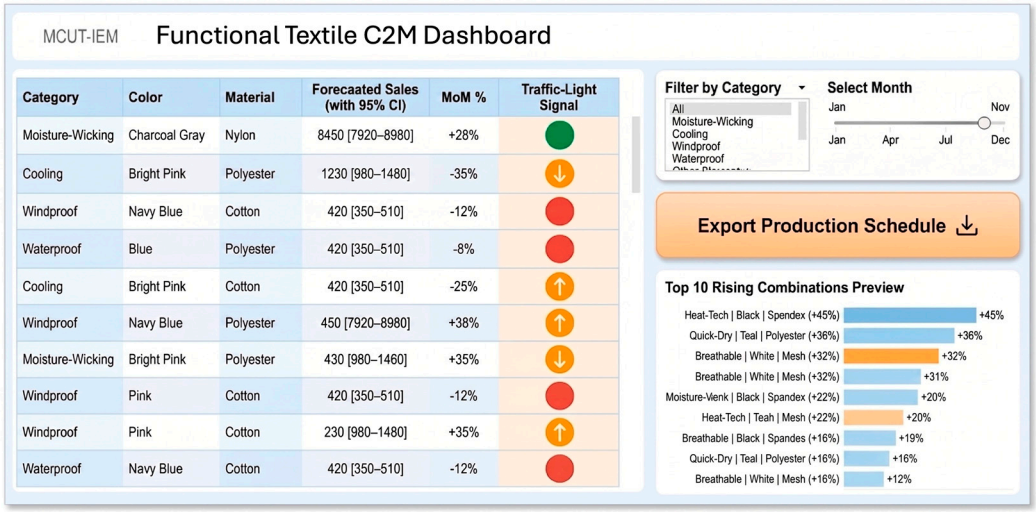

Figure 6. Human-centric C2M dashboard featuring traffic-light signals, attribute-level drill-down, and one-click export for legacy ERP integration.

The decision logic is based on the relative position of the predicted expected sales $E[\hat{S}_i]$ within the category-specific empirical distribution (where P_k denotes the k-th percentile), in conjunction with the month-over-month (MoM) growth rate (Δ_{MoM}). High- and low-demand regimes are identified using percentile-based thresholds, whereas short-term momentum is captured by growth signals. The resulting decision regions are mutually exclusive and collectively exhaustive, ensuring consistent and interpretable signal assignment. The formal specification of the decision rules is provided in Algorithm 1 as follows:

Algorithm 1. Traffic-Light Signal Assignment for Human-in-the-Loop Production Planning

Input:

- Predicted expected sales for item i: $E[\hat{S}_i]$
- Category-specific empirical sales distribution percentiles:
$$\{P_5, P_{20}, P_{80}, P_{95}\}$$
- Month-over-Month growth rate: Δ_{MoM}

Output:

- Decision signal $\in$ { Green, Rising Yellow, Falling Yellow, Red, Neutral}

Steps:

1. If $E[\hat{S}_i] \geq P_{95}$, assign Green (Immediate Procurement).
2. Else if $P_{80} \leq E[\hat{S}_i] < P_{95}$ and $\Delta_{MoM} > +20\%$, assign Rising Yellow (Emerging Trend).
3. Else if $P_5 < E[\hat{S}_i] \leq P_{20}$ and $\Delta_{MoM} < -20\%$, assign Falling Yellow (Obsolescence Risk).
4. Else if $E[\hat{S}_i] \leq P_5$, assign Red (Clearance/Halt).
5. Else, assign Neutral (Monitoring only).

Traffic-light signals translate probabilistic demand forecasts into interpretable managerial actions by mapping forecast uncertainty and demand dynamics into structured decision regions. Percentile-based thresholds ensure robustness across product categories with heterogeneous demand scales, and the incorporation of short-term month-over-month growth captures dynamic demand transitions. The resulting decision regions are mutually exclusive and collectively exhaustive, thereby preventing ambiguous or conflicting signals.

From an operational perspective, products residing in the upper tail of the empirical demand distribution are flagged as high-priority items, triggering immediate raw material reservations to mitigate stockout risk. Conversely, products in the lower tail are identified as candidates for production suspension or inventory clearance. Intermediate demand regimes are further differentiated by short-term demand momentum, enabling planners to distinguish emerging trends from obsolescence risks. Products that do not exhibit pronounced demand signals or significant growth or decline are classified as neutral and monitored without immediate intervention.

The deployed dashboard was explicitly designed to enhance cognitive efficiency. Visualizing 95% credible intervals alongside point forecasts enables non-technical managers to intuitively assess forecast uncertainty without requiring advanced statistical training. Furthermore, the system supports dynamic, multi-dimensional filtering (e.g., isolating nylon-based products during summer months) and one-click export of ranked production schedules in CSV format, facilitating low-friction integration with legacy ERP systems commonly used by small and medium-sized enterprises.

A distinguishing feature of this framework is its ability to operationalize Bayesian uncertainty quantification within a user-friendly environment. Instead of automating

production decisions, the traffic-light interface provides structured decision boundaries that support risk-informed human judgment.

4. Analysis Results

4.1. Model Benchmarking and Selection: Ensuring Algorithmic Robustness

To identify a robust forecasting engine for the proposed C2M framework, five candidate algorithms were evaluated using identical stratified train–validation splits (80%/20%). The benchmarking process comprised two phases: an initial screening to assess architectural suitability under high-cardinality demand signals, followed by intensive optimization of the shortlisted models.

- Phase 1: Initial screening and generalization analysis.

During the initial screening using default hyperparameters, standard gradient boosting architectures demonstrated pronounced discrepancies between training and validation performance. While XGBoost and Gradient-Boosted Trees achieved high goodness-of-fit on the training data ($R^2 > 0.93$), their validation performance deteriorated substantially (mean validation $R^2 = 0.353$ and 0.550, respectively), indicating limited generalization under sparse, high-cardinality attribute regimes. Consequently, both models were excluded from subsequent optimization stages.

In contrast, Neural Boosted Trees and Bootstrap Forest exhibited more stable generalization, maintaining consistently narrower train–validation gaps and achieving higher validation performance (mean validation $R^2 = 0.802$ and 0.753, respectively). These results motivated their selection for further optimization.

- Phase 2: Hyperparameter optimization and final model selection.

The two shortlisted models underwent intensive hyperparameter optimization using a 50-trial Tree-structured Parzen Estimator (TPE) search, combined with 5-fold stratified cross-validation. Model performance was evaluated using cross-validated R^2, root mean squared error (RMSE), and prediction interval (PI) coverage. Across all functional categories, Neural Boosted Trees consistently outperformed Bootstrap Forest. A paired t-test of fold-level R^2 scores indicated statistically significant differences ($p < 0.01$). Consequently, Neural Boosted Trees were selected for final evaluation, and their cross-validated performance is reported in the Results section.

The final 5-fold cross-validated performance of the selected Neural Boosted Tree model is summarized in Table 4. The model achieved strong predictive accuracy across all functional apparel categories, with R^2 values ranging from 0.854 to 0.980. Performance was highest for moisture-wicking and cooling apparel, reflecting more stable and frequent demand signals in these categories. Importantly, the model maintained consistent uncertainty calibration, with prediction interval coverage close to the nominal 95% level across all categories. This balance between point accuracy and uncertainty reliability supports the suitability of the proposed approach for upstream production planning, where both forecast precision and risk awareness are required.

Table 4. Final 5-fold cross-validated performance of the selected Neural Boosted Tree model.

Category	R^2	RMSE	95% PI Coverage
Moisture-Wicking	0.980	12.4	94.8%
Cooling	0.986	11.1	95.2%
Thermal	0.862	38.2	93.1%
Windproof/Waterproof	0.854	13.8	94.5%
Overall	0.921	18.9	94.4%

The selected Neural Boosted Trees model achieved an overall R^2 of 0.921. Crucially, it demonstrated a 95% Prediction Interval (PI) coverage of 94.4%, indicating excellent probabilistic calibration. This metric is vital for the downstream traffic-light system, as it ensures that the communicated "risk" to human decision-makers is statistically accurate. These results support the hypothesis that the proposed hybrid architecture, which combines neural representations of high-order categorical interactions with tree-based residual correction, is well-suited to noisy, high-dimensional tabular data derived from publicly available data streams.

4.2. Temporal Generalization and Resilience to Regime Shifts

To evaluate the system's robustness to temporal concept drift, the final model trained on data from 2021–2022 was prospectively assessed using a hold-out dataset spanning January to June 2023 (N = 18,426 new observations). This period constitutes a challenging regime transition from pandemic-influenced consumption toward post-recovery market normalization.

Despite concurrent volatility in raw material costs and consumer sentiment, the aggregate hold-out evaluation yielded an R^2 of 0.907, corresponding to a modest performance degradation of approximately 1.5% relative to cross-validation results. This limited decline indicates that the Neural Boosted Tree model maintained stable predictive capacity under a temporal regime shift, capturing demand-relevant attribute semantics without overfitting to transient, shock-driven patterns observed during the pandemic period. Category-specific results (Table 5) further reveal interpretable seasonal behaviors aligned with the physical usage characteristics of apparel products, such as peak sensitivity in thermal categories during winter months. These findings support the suitability of the proposed framework as a robust, human-centric decision-support tool that provides reliable guidance during periods of market transition, rather than solely under stationary demand conditions.

Table 5. Performance data from the January 2023 holdout data (Winter Stress Test).

Category	R^2	RMSE	Interpretation
Moisture-Wicking	0.535	174.6	Moderate seasonality impact
Cooling	−0.412	432.8	Zero-inflation artifact (Winter season)
Thermal	0.901	1464.4	High fidelity during peak demand
Windproof/Waterproof	0.596	234.1	Stable baseline performance

- Thermal apparel demonstrated the highest R^2 value (0.901) during the winter test period (January–June), indicating strong alignment between model forecasts and peak-season demand dynamics when forecasting accuracy is most critical.
- Cooling apparel exhibited a negative R^2 value (−0.412) in January. This result reflects a seasonal zero-demand effect rather than a deficiency of the forecasting model. During winter months, the true demand for cooling fabrics approaches zero and exhibits minimal variance, a condition under which standard regression-based metrics such as R^2 become unstable and may yield negative values despite small absolute prediction errors. In practical terms, this category represents a structurally inactive demand regime rather than a forecasting failure. Accordingly, winter cooling apparel forecasts were treated as low-priority signals in the decision-support system and did not trigger production actions, ensuring that the observed metric anomaly did not affect operational planning or system robustness.

This edge case highlights the importance of incorporating human-in-the-loop decision support (Section 3.4.2). Fully automated, opaque forecasting systems may generate misleading alerts when performance metrics are affected by seasonal or statistical artifacts.

Conversely, the proposed traffic-light interface enables human planners to contextualize model outputs and filter out seasonally irrelevant signals. This design reflects the principles of Industry 5.0, wherein artificial intelligence is intended to augment, rather than replace, human judgment in complex manufacturing environments.

4.3. Longitudinal Field Validation: Operational Impact and Environmental, Social, and Governance (ESG) Implications

In July 2023, the full C2M framework was deployed at a partner textile SME in northern Taiwan (annual production capacity of approximately 1.2 million meters) for a 12-month longitudinal field validation. The objective was to evaluate the operational implications of transitioning from a reactive, push-based order planning paradigm to a pull-based C2M intelligence framework. Performance outcomes were assessed through a before–and–after comparison, with the 12-month period preceding deployment serving as the baseline.

- Financial resilience (inventory optimization)

The average monthly inventory holding value decreased by approximately 28%, from NT$18.4 million in the pre-deployment period to NT$13.2 million in the post-deployment period. Inventory value was calculated based on month-end stock levels recorded in the firm's enterprise resource planning system. The release of working capital tied to slow-moving and obsolete inventory directly strengthened the firm's short-term financial resilience under demand volatility.

- Sustainable manufacturing (process efficiency and resource use)

Following deployment, the number of dye lot changeovers decreased by 31%. Incorporating attribute-level demand forecasts into production planning enabled the consolidation of small and fragmented orders into forecast-guided batch sizes. Reduced changeovers decreased the frequency of machine washdowns, which are resource-intensive operations in textile dyeing. This operational shift reduced water, energy, and chemical consumption, aligning production planning decisions with environmental sustainability objectives.

- Operational agility (capacity utilization)

The effective capacity utilization increased from an average of 68% during the baseline period to 84% following deployment. This improvement was primarily due to reduced idle time, stemming from frequent setup changes and unanticipated material shortages. Increased utilization allows the SME to accommodate shorter lead times or rush orders that were previously infeasible under the push-based planning regime.

In addition to quantitative indicators, system-level operational traces provided context for the observed performance changes. Analysis of dashboard access logs, planning-cycle timestamps, and event records of schedule revisions and order releases demonstrated that the traffic light dashboard reduced information latency between market-demand signals and manufacturing actions by approximately six to eight weeks. This reduction mitigated the amplification effects typically associated with downstream order volatility, thereby enhancing responsiveness and coordination in alignment with Industry 5.0 principles.

5. Discussion

5.1. Theoretical Contributions

This study advances the literature on fashion and textile supply chains by addressing persistent gaps in upstream demand sensing, semantic translation, and human-centric decision support.

First, this research extends the customer-to-manufacturer (C2M) paradigm beyond predominantly downstream, retail-centric applications to upstream, component-level man-

ufacturing contexts. Previous C2M studies have primarily examined finished-product forecasting and retailer responsiveness [6]. From an information-processing perspective [35], upstream small and medium-sized enterprises (SMEs) experience the highest information-uncertainty load in the textile value chain but receive the least timely market signals. This structural asymmetry is directly addressed by the proposed framework. This study demonstrates that consumer-generated signals can be systematically translated into attribute-level intelligence, specifically color, material, and function, which are directly actionable for upstream dye-batch scheduling. The observed 6–8-week reduction in information latency during field deployment aligns with Lee et al.'s [4] theoretical prediction that shared demand information can substantially reduce bullwhip amplification upstream. This result provides empirical corroboration in an SME context that was previously absent from the C2M literature and addresses a key limitation regarding the applicability of C2M to upstream SMEs operating under high product variety and fragmented orders.

Second, this research advances the study of Industry 5.0 by integrating human-centric decision support into digital forecasting, thereby contributing to the dynamic capabilities literature as applied to SME digitalization [36,37]. In contrast to Industry 4.0, which emphasizes black-box automation, Industry 5.0 prioritizes resilience and the ongoing role of human participation [3]. The findings indicate that resilience does not require full automation. Instead, the artificial intelligence system functions as a noise filter, processing high-frequency data so human planners can focus on high-value signals. By utilizing probabilistic forecasts to address approximately 80% of data noise, the system enables human operators to manage the remaining 20% of complex exceptions. This approach is exemplified by the 'winter stress test,' in which human judgment correctly identified a zero-demand artifact in cooling fabrics that standard metrics might misinterpret.

Third, this research makes a methodological contribution by addressing high-cardinality categorical data within demand signals. The results extend the empirical findings of Shwartz-Ziv and Armon [34], who argued that deep learning does not universally outperform tree-based methods on tabular data. This study identifies a boundary condition of extreme categorical cardinality combined with semantic sparsity, where hybrid neural-boosting architectures outperform both pure deep learning and tree-based methods. Furthermore, validating 'Comment Count' as a high-fidelity proxy for sales (Pearson's $r = 0.954$) expands the methodological toolkit for demand sensing in contexts where direct sales observations are unavailable.

5.2. Practical and Managerial Contributions

This study provides targeted, actionable insights for five practitioner groups: production planning teams, supply chain and procurement managers, SME owners and financial decision-makers, sustainability officers, and industrial policymakers. Collectively, these implications demonstrate that the proposed framework offers not only academic value but also serves as a practical intelligence tool for resource-constrained upstream manufacturers.

For production planning teams, the traffic-light dashboard offers an immediately deployable decision-support interface that does not require advanced data science expertise. Production planners may incorporate monthly C2M intelligence cycles using publicly available Shopee Taiwan consumer data, with data refreshes scheduled to align with Northern Hemisphere athletic apparel purchasing cycles (March–May and September–November). Evidence from the field shows that traditional order-lagged planning results in demand-signal latency of 6–8 weeks. Implementing weekly data refresh cycles reduces this latency to near-real-time, enabling proactive dye-batch scheduling ahead of peak seasonal demand. Planners should also identify color–material combinations that maintain green-signal status over two or more consecutive monthly cycles, as these indicate structurally stable demand

anchors suitable for longer-term capacity commitments and reduced emergency setups. The system's one-click CSV export is compatible with legacy ERP environments prevalent in Taiwanese mid-tier dyeing firms, removing the need for custom IT integration.

For supply chain and procurement managers, the probabilistic forecast outputs—specifically the 95% credible intervals generated by the Bayesian Monte Carlo engine—offer a quantitative basis for supplier negotiations. Procurement teams can use sustained green-signal attribute clusters to justify consolidating minimum order quantities with upstream yarn and chemical suppliers, thereby reducing per-batch procurement costs and inventory exposure. The documented 31% reduction in dye-lot changeovers at the partner SME demonstrates that forecast-guided batch consolidation is operationally feasible within a single planning cycle. Each avoided changeover results in an estimated reduction of 800–1200 L of process water, a metric that can be directly incorporated into procurement contracts specifying environmental performance standards. Procurement managers at firms supplying European sports brands may also use this evidence to negotiate preferential terms under sustainability-linked purchasing frameworks, as reduced changeover rates correspond to verifiable Scope 3 supply chain emission reductions.

For SME owners and financial decision-makers, the total implementation cost at the partner firm was less than NT$150,000 (approximately USD 4700), achieved entirely through open-source software and publicly available e-commerce data. For SMEs considering digital transformation under capital constraints, the framework provides a phased adoption pathway: starting with data acquisition and demand sensing, and gradually integrating the semantic ontology and dashboard as operational familiarity increases. The increase in capacity utilization from 68% to 84% post-deployment—a 16-percentage-point improvement on a 1.2-million-meter annual production base—results in significant incremental revenue without additional capital expenditure. Additionally, the 28% reduction in average monthly inventory holding value (from NT$18.4 million to NT$13.2 million) released approximately NT$5.2 million in working capital per month, thereby strengthening the firm's liquidity position amid demand volatility. This cost–benefit profile indicates that payback periods for similar implementations may be achievable within a single fiscal quarter.

For sustainability officers and ESG reporting teams, the framework produces audit-ready operational data—including changeover frequency trajectories, inventory turnover records, and capacity utilization logs—suitable for buyer-facing ESG reporting under internationally recognized standards. The 31% reduction in dye-lot changeovers supports disclosure under ISO 14001 (Environmental Management Systems) and Science-Based Targets (SBT) frameworks, which are increasingly required by European brand clients. Sustainability officers should establish pre-deployment baselines for water, energy, and chemical consumption per changeover event to enable rigorous pre- and post-implementation impact quantification in accordance with GRI Standard 303 (Water and Effluents) and GRI 302 (Energy). The system's operational logs also provide verifiable evidence for Scope 3 supply chain emission disclosures, which are increasingly subject to third-party audit requirements under EU sustainability reporting regulations (CSRD). For firms seeking competitive advantage in global sustainable sourcing programs, this quantifiable evidence base constitutes a differentiated capability not achievable through non-data-driven production planning approaches.

For industrial policymakers and sector associations, this study demonstrates that the digital divide between upstream SMEs and large-brand supply chain ecosystems can be significantly reduced through the use of open-source tools and publicly available data. Government agencies, such as the Taiwan Textile Research Institute (TTRI) and the Industrial Development Bureau, are encouraged to develop shared C2M intelligence portals that aggregate e-commerce platform data across participating SME clusters, thereby lowering

per-firm implementation costs through collective data governance. Regional industrial associations may adopt the framework's architecture as a shared intelligence service for member firms, enabling collective downstream demand visibility without requiring each SME to independently establish data acquisition pipelines. These policy instruments would directly support the Ministry of Economic Affairs' Smart Machinery and Sustainable Manufacturing Initiative, positioning Taiwan's upstream textile sector as a model for Industry 5.0-aligned SME digitalization in export-oriented manufacturing economies.

Practitioners implementing this framework should consider two operational boundary conditions. First, the framework was validated for the upstream dyeing-and-finishing segment; application to downstream cut-and-sew or accessories manufacturing would require recalibration of the semantic ontology and proxy validation procedures. Second, the traffic-light percentile thresholds (P_5, P_{20}, P_{80}, P_{95}) are category-specific and should be recalibrated annually to reflect changes in platform-level demand structure, especially after significant algorithmic updates or post-pandemic demand normalization.

6. Conclusions

This study presents a validated, scalable pathway to enhance supply chain resilience in upstream textile small and medium-sized enterprises (SMEs) by integrating public e-commerce data into a human-centric Industry 5.0 framework. Through the development and field validation of a customer-to-manufacturer (C2M) decision-support system, this research demonstrates that unstructured consumer signals can be systematically converted into detailed, attribute-level intelligence regarding color, material, and functional properties relevant to upstream dyeing and finishing operations. Longitudinal industrial deployment results show that shifting from reactive, order-lagged planning to proactive, demand-informed decision-making yields substantial operational and ESG-aligned benefits. The implementation led to a 28% reduction in inventory value, a 31% decrease in dye-lot changeovers, and a 16% increase in capacity utilization. Rather than supplanting human expertise, the proposed framework acts as an intelligent noise filter, reducing information latency by approximately six to eight weeks and enabling shop-floor planners to prioritize significant market shifts. This research provides empirical support for Industry 5.0 principles in resource-constrained settings, illustrating how affordable digital intelligence can strengthen the resilience and sustainability of upstream manufacturing.

For upstream Taiwanese textile SMEs, particularly those engaged in dyeing and finishing operations supplying functional sportswear and outdoor apparel fabrics to global brands, this study provides four specific and actionable recommendations. First, production planning teams are advised to implement a monthly C2M intelligence cycle based on Shopee Taiwan consumer data, aligned with Northern Hemisphere seasonal athletic apparel purchasing patterns. Field evidence indicates that demand-signal latency peaks at 6 to 8 weeks under conventional order-lagged planning; adopting weekly data refresh cycles can reduce this latency to near real-time, enabling proactive dye-batch scheduling ahead of peak athletic seasons (March to May and September to November). Second, SME production managers can use the traffic-light dashboard to identify color–material combinations that sustain green-signal status over multiple months and use this information to negotiate minimum order quantities with upstream yarn and chemical suppliers, thereby consolidating dye batches. The observed 31% reduction in dye-lot changeovers at the partner SME decreased machine washdown frequency, with each eliminated changeover saving an estimated 800 to 1200 L of process water, directly contributing to quantifiable ESG outcomes aligned with ISO 14001 documentation requirements increasingly required by European brand clients, such as Nike and Patagonia. Third, the framework's operational data, including changeover frequency, inventory turnover, and capacity utilization trajectories,

provide audit-ready sustainability metrics suitable for buyer-facing ESG reporting, thereby enhancing the competitive positioning of Taiwan's functional-fabric manufacturers in global sustainable sourcing programs. Fourth, the total implementation cost at the partner SME was below NT$150,000 (approximately USD 4700), achieved entirely with open-source software and publicly available data. This demonstrates that the transition from reactive order-following to proactive demand-informed planning is financially accessible to mid-tier dyeing firms without the need for proprietary IT investment.

Despite these contributions, this study has certain limitations. First, empirical validation was conducted on a single e-commerce platform within a localized market, which may constrain generalizability across different cultural, regulatory, or platform-specific contexts. Second, although consumer engagement metrics such as comment counts served as high-fidelity proxies for actual demand ($r = 0.954$), these remained indirect indicators and were subject to platform-specific algorithmic variations and promotional influences.

Future research should incorporate multi-platform data streams, including social media and global cross-border marketplaces, to improve the robustness and scope of demand sensing. Additionally, advancements in natural language processing and large language models may enable a more detailed extraction of functional nuances and granular consumer sentiment from unstructured text. The integration of online learning mechanisms could allow forecasting models to adapt dynamically to rapid regime shifts and non-stationary fashion cycles. For studies focusing on Taiwan's upstream textile sector, it is important to assess whether the seasonal demand-signal patterns identified in this research, specifically the March to May and September to November athletic apparel procurement cycles, apply to other functional-fabric product segments, such as technical outerwear and medical-grade performance textiles, which are expanding export categories for Taiwanese suppliers. Finally, future studies should quantify environmental impacts with greater specificity by directly linking reduced setup frequencies to reductions in water, energy, and carbon footprints, with particular emphasis on the dyeing and finishing segment, given its significant contribution to Taiwan's textile sector's water consumption. This approach would further advance the sustainability objectives of Industry 5.0.

Author Contributions: Conceptualization, C.-C.W., Y.-T.H. and H.-Y.K.; methodology, C.-C.W., Y.-T.H. and H.-Y.K.; software, Y.-T.H.; validation, C.-C.W., Y.-T.H. and H.-Y.K.; formal analysis, Y.-T.H. and H.-Y.K.; resources, C.-C.W.; data curation, Y.-T.H. and H.-Y.K.; writing—original draft preparation, Y.-T.H. and H.-Y.K.; writing—review and editing, C.-C.W.; visualization, C.-C.W.; supervision, C.-C.W.; funding acquisition, C.-C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the Ministry of Science and Technology Council of Taiwan (111-2813-C-131-008-E).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Access to these data is restricted in accordance with the platform terms. The derived features and processing logic can be obtained from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fisher, M.L. What is the right supply chain for your product? *Harv. Bus. Rev.* **1997**, *75*, 105–116.
2. 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*, 530–535. [CrossRef]

3. Breque, M.; De Nul, L.; Petridis, A. *Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry*; Publications Office of the European Union: Luxembourg, 2021; KI-BD-20-021-EN-N.
4. Lee, H.L.; Padmanabhan, V.; Whang, S. Information distortion in a supply chain: The bullwhip effect. *Manag. Sci.* **1997**, *43*, 546–558. [CrossRef]
5. Ghobakhloo, M. Industry 5.0 and the future of manufacturing: A systematic review and research agenda. *Technol. Forecast. Soc. Chang.* **2023**, *190*, 122398.
6. He, B.; Mirchandani, P.; Yang, G. Offering custom products using a C2M model: Collaborating with an e-commerce platform. *Int. J. Prod. Econ.* **2023**, *262*, 108918. [CrossRef]
7. Swaminathan, K.; Venkitasubramony, R. Demand forecasting for fashion products: A systematic review. *Int. J. Forecast.* **2024**, *40*, 247–267. [CrossRef]
8. Christopher, M.; Lowson, R.; Peck, H. Creating agile supply chains in the fashion industry. *Int. J. Retail. Distrib. Manag.* **2004**, *32*, 367–376. [CrossRef]
9. Kalaoglu, Ö.İ.; Akyuz, E.S.; Ecemiş, S.; Eryuruk, S.H.; Sümen, H.; Kalaoglu, F. Retail demand forecasting in clothing industry. *Text. Appar.* **2015**, *25*, 172–178.
10. Choi, T.M.; Wallace, S.W.; Wang, Y. Big data analytics in operations management. *Prod. Oper. Manag.* **2018**, *27*, 1868–1883. [CrossRef]
11. Cachon, G.P.; Fisher, M. Supply chain inventory management and the value of shared information. *Manag. Sci.* **2000**, *46*, 1032–1048. [CrossRef]
12. Chowdhury, A.R.; Paul, R.; Rozony, F.Z. A systematic review of demand forecasting models for retail e-commerce enhancing accuracy in inventory and delivery planning. *Int. J. Sci. Interdiscip. Res.* **2025**, *6*, 1–27. [CrossRef]
13. Grooss, O.F. Digitalization of maintenance activities in small and medium-sized enterprises: A conceptual framework. *Comput. Ind.* **2024**, *154*, 104039. [CrossRef]
14. Yilmaz, G.; Salter, L.; McFarlane, D.; Schönfuß, B. Low-cost (Shoestring) digital solution areas for enabling digitalisation in construction SMEs. *Comput. Ind.* **2023**, *150*, 103941. [CrossRef]
15. Wang, W.; Wang, Y.; Chen, Y.; Deveci, M.; Kadry, S.; Pedrycz, W. Analyzing the barriers to resilience supply chain adoption in the food industry using hybrid interval-valued fermatean fuzzy PROMETHEE-II model. *J. Ind. Inf. Integr.* **2024**, *40*, 100614. [CrossRef]
16. Shubber, M.S.; Mohammed, M.T.; Qahtan, S.; Ibrahim, H.A.; Mourad, N.; Zaidan, A.A.; Zaidan, B.B.; Deveci, M.; Pamucar, D.; Wu, P. Pythagorean fuzzy rough decision-based approach for developing supply chain resilience framework in the face of unforeseen disruptions. *J. Ind. Inf. Integr.* **2025**, *45*, 100837. [CrossRef]
17. Pistolesi, F.; Baldassini, M.; Lazzerini, B. A human-centric system combining smartwatch and LiDAR data to assess the risk of musculoskeletal disorders and improve ergonomics of Industry 5.0 manufacturing workers. *Comput. Ind.* **2024**, *155*, 104042. [CrossRef]
18. Wang, T.C.; Guo, R.S.; Chen, C. An integrated data-driven procedure for product specification recommendation optimization with LDA-LightGBM and QFD. *Sustainability* **2023**, *15*, 13642. [CrossRef]
19. Ren, S.; Chan, H.L.; Siqin, T. Demand forecasting in retail operations for fashionable products: Methods, practices, and real case study. *Ann. Oper. Res.* **2020**, *291*, 761–777. [CrossRef]
20. Niu, T.; Chen, X. Consumer-to-manufacturer: Supply chain and hybrid contracts. *Manag. Decis.* **2025**, 1–24. [CrossRef]
21. Satish, A. Machine learning-driven demand forecasting: A comparative analysis of advanced techniques and real-time integration. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **2024**, *10*, 1352–1361. [CrossRef]
22. Tortorella, G.L.; Powell, D.; Liu, L.; Filho, M.G.; Antony, J.; Hines, P.; Nascimento, D.L.D.M. How has social media been affecting problem-solving in organizations undergoing Lean Production implementation? A multi-case study. *J. Ind. Inf. Integr.* **2023**, *35*, 100515. [CrossRef]
23. Sestino, A.; Prete, M.I.; Piper, L.; Guido, G. Internet of Things and big data as enablers for business digitalization strategies. *Technovation* **2020**, *98*, 102173. [CrossRef]
24. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794.
25. Villegas, M.A.; Pedregal, D.J.; Trapero, J.R. A support vector machine for model selection in demand forecasting applications. *Comput. Ind. Eng.* **2018**, *121*, 1–7. [CrossRef]
26. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [CrossRef]
27. Baryannis, G.; Dani, S.; Antoniou, G. Predictive analytics and artificial intelligence in supply chain management: Review and implications for the future. *Comput. Ind. Eng.* **2019**, *137*, 106024.

28. Lim, J.; Patel, S.; Evans, A.; Pimley, J.; Li, Y.; Kovalenko, I. Enhancing human-robot collaborative assembly in manufacturing systems using large language models. In *Proceedings of the IEEE 20th International Conference on Automation Science and Engineering, Bari, Italy, 28 August–1 September 2024*; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2024; pp. 2581–2587.
29. Mao, Z.; Zhang, J.; Sun, Y.; Fang, K.; Huang, D. Balancing parallel assembly lines with human-robot collaboration: Problem definition, mathematical model and tabu search approach. *Int. J. Prod. Res.* **2025**, *63*, 51–85.
30. Kim, M.; Shim, J.Y.; Lim, S.; Lee, H.; Kwon, S.C.; Hong, S.; Ryu, S. Reduction of greenhouse gas emissions by optimizing the textile dyeing process using digital twin technology. *Fash. Text.* **2024**, *11*, 17. [CrossRef]
31. Wieland, A.; Durach, C.F. Two perspectives on supply chain resilience. *J. Bus. Logist.* **2021**, *42*, 315–322. [CrossRef]
32. Choi, T.M.; Kumar, S.; Yue, X.; Chan, H.L. Disruptive technologies and operations management in the Industry 4.0 era and beyond. *Prod. Oper. Manag.* **2022**, *31*, 9–31. [CrossRef]
33. Bag, S.; Gupta, S.; Kumar, S. Industry 4.0 adoption and 10R advance manufacturing capabilities for sustainable development. *Int. J. Prod. Econ.* **2021**, *231*, 107844. [CrossRef]
34. Shwartz-Ziv, R.; Armon, A. Tabular data: Deep learning is not all you need. *Inf. Fusion* **2022**, *81*, 84–90. [CrossRef]
35. Galbraith, J.R. Organization design: An information processing view. *Interfaces* **1974**, *4*, 28–36. [CrossRef]
36. Teece, D.J. Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strateg. Manag. J.* **2007**, *28*, 1319–1350. [CrossRef]
37. Ghobakhloo, M.; Iranmanesh, M. Digital transformation success under Industry 4.0: A strategic guideline for manufacturing SMEs. *J. Manuf. Technol. Manag.* **2021**, *32*, 1533–1556. [CrossRef]

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