Opsi
Vol. 19, No. 1, June 2026
p-ISSN 1693-2102
e-ISSN 2686-2352

Proactive quality control in cotton yarn manufacturing using random forest defect forecasting and CTQ analysis

Adhie Tri Wahyudi [1*], **Mayang Eka Putri Prastiwi** [1], **Erni Suparti** [1], **Muhammad Rizki** [2]
[1]Department of Industrial Engineering, Universitas Setia Budi, Mojosongo, Surakarta 57127, Indonesia
[2]Department of Industrial Management, National Taiwan University of Science and Technology, Taipei City 106335, Taiwan

*Corresponding Author: adhie@setiabudi.ac.id

Article history:
Received: 19 April 2026
Revised: 11 June 2026
Accepted: 29 July 2026
Published: 30 July 2026

Keywords:
Quality control
Defect forecasting
CTQ
Cotton yarn
Random forest

ABSTRACT
Quality control in cotton yarn manufacturing is affected by the daily variation in the occurrence of defects and the late detection of abnormal events at the packing stage. This study proposes a decision-support workflow that integrates Random Forest regression for daily defect forecasting with Critical-to-Quality (CTQ) analysis to enable proactive quality control in an industrial case. The model is trained and validated with time ordered evaluation using historical production volume and defect counts, and the performance is reported using MAE, RMSE and sMAPE. Results indicate that the model learns to capture routine up-down defect patterns during the training period, but prediction accuracy is degraded on test days with extreme defect spikes. However, despite this limitation, the forecast outputs provide an early warning signal to prioritize inspection and corrective actions on high risk days. The predicted risk levels are then further linked to CTQ mapping to identify dominant defect types and convert model outputs to actionable defect-specific control recommendations. The proposed integration of machine-learning forecasting with shop-floor quality planning enables more timely and targeted quality interventions in cotton yarn packaging operations.

DOI:
https://doi.org/10.31315/opsi.v19i1.16757

1. INTRODUCTION

The Textile and Textile Product (TPT) industry is one of the strategic manufacturing sectors in Indonesia. It contributes to the non-oil and gas processing industry GDP by 6.38 percent and employs around 3.6 million workers [1]. As of mid-2022, the share of TPT commodities in Indonesia's total national export reached 5.51 or equivalent to USD 6.08 billion [1]. However, the industry faces increasing competitive pressure. The export value of Indonesia's TPT continues to decline from 2023, whereas import volumes have increased in 2024 [2]. This trend indicates that the domestic textile industry is under increasing competition in the domestic and international markets. In this perspective, product quality is not just an operational standard, but a crucial competitive imperative. Defects have a significant impact on a cost efficiency, delivery reliability, and brand reputation which are crucial for TPT manufacturers to remain competitive in the market.

The object of this study is cotton yarn. The majority of quality issues occur during the packing process, including crossed yarn, soiling, color stains, bad winding, loose packages, mismatched cones, debris, and broken yarn. Such defects decrease quality standards compliance and increase scrap, rework, and inspection workload. By August 2025, the defect rate was about 2% of total yarn production or about 20,000 Defects Per Million Opportunities (DPMO). This figure is lower than the benchmark reported in similar cotton yarn quality studies where DPMO of around 7,786 at sigma level of 4 has been documented [3], and considerably below the Six Sigma world-class threshold of 3,400 DPMO [3], [4]. This gap indicates that the quality level of current process is such that a systematic intervention is justified. A related study by Harianto et al. [5] stated a potential loss of around IDR 75.6 million per month caused by the tangled yarn due to the ring-touch phenomenon. This example highlights the economic implications of yarn defects. Previous studies have shown that higher defect rates can increase rework, scrap costs and disrupt the production output stability due to the additional time and resources required to handle nonconforming products [5], [6]. Such instability may adversely affect production scheduling and resource utilization, including labor allocation and machine availability, particularly in labor-intensive industries such as TPT [5]. Furthermore, frequent quality deviations often require additional inspections, repacking activities, and process adjustments, thereby increasing operational complexity [6]. These conditions may ultimately reduce equipment effectiveness and limit a factory's ability to respond quickly to fluctuations in customer demand [5], [6].

Furthermore the fluctuation of daily production reduces the planning reliability across upstream and downstream processes, including warehousing and logistics. Thus, the on-time delivery performance may deteriorate, which negatively affects customer satisfaction and contractual compliance. Sahoo [6] has pointed out that rework exceeding 7% can reduce productivity and lead to the waste of resources. Rohmat et al. [7] also described that product quality influences customer satisfaction, customer loyalty, and brand trust.

Previous studies have attempted to reduce yarn defects using process improvement methods. Wahyudin et al. [8] and Perwitasari et al. [9] applied Six Sigma using the DMAIC methodology. DMAIC refers to the acronym for Define, Measure, Analyze, Improve, and Control. This methodology originates from the Six Sigma framework and is used to identify, analyze, and eliminate defects or inefficiencies in business processes. They reported greater sigma levels and decreased DPMO. This approach is structured, useful, practical, and well-organized. However, it is mostly post-hoc. It depends on historical events and does not directly provide a daily early-warning mechanism. A purely reactive quality control approach is often insufficient. Defects are only detected after they happen and do not prevent abnormal defect days. This limitation becomes critical during extreme defect spikes. A more proactive strategy is required by the industry [6], [10]. This strategy should identify high-risk days early on. Then, preventive actions can be taken before losses accumulate. While Six Sigma provides a structured framework for defect reduction, its retrospective nature limits its ability to anticipate defect occurrences before they happen. Addressing this limitation requires a shift toward data-driven predictive approaches capable of generating daily risk forecasts.

Predictive approaches in quality control have shown significant growth across manufacturing research. Iskandar et al. [10] developed an Artificial Neural Network (ANN) to predict end breakage defects. They used technical variables such as rotor speed and trash content. Choi et al. [11] also proposed a defect prediction based on a data-driven approach. This shift toward predictive methods has important implications for quality control practices, as it moves the focus from outcome-based inspection to risk-based forecasting before products are processed or packed. As a result, quality control becomes less reactive and more proactive, enabling manufacturers to plan inspections, corrective actions, and resource allocation in a more timely and targeted manner.

In recent manufacturing literature, machine learning (ML) is widely used for prediction, classification, and anomaly detection. Ensemble learning is particularly relevant. Random Forest (RF) is one of the popular ensemble methods that can model nonlinear relationships and is resilient to noise or outliers. Breiman's and subsequent publications [12], [13] discussed the foundations and advantages of RF. The Engineering Journal also reports manufacturing decision-support applications using machine learning. Junpradub and Asawarungsaengkul [14] applied ML techniques to classify hot-rolled steel sheet products and demonstrated that RF delivered superior performance. They also emphasized its potential to reduce trial-run costs, reinforcing the relevance of machine learning for industrial decision-making. Given its robustness to noisy operational data, ability to model complex nonlinear relationships, and interpretability through feature importance measures, RF represents an effective approach for supporting quality control decisions in manufacturing environments [13], [15].

While Iskandar et al. [10] predicted end breakage defects using ANN with sensor-based process variables, Igbokwe et al. [16] demonstrated ML-based defect classification in a manufacturing SME, and Choi et al. [11] developed defect prediction models from spinning shop floor data, these studies focus primarily on predictive accuracy without translating outputs into operational early-warning mechanisms or actionable quality control steps. The integration of daily defect forecasts with CTQ-based inspection prioritization remains unaddressed in the literature. This operational translation gap represents the primary contribution of the present study. This study therefore aims to develop a RF-based daily defect prediction model integrated with CTQ analysis to support proactive, actionable quality control at the packing stage of cotton yarn manufacturing. The prediction outputs are not merely treated as numerical forecasts; instead, they serve as early-warning indicators for identifying high-risk defect days. However, prediction alone is insufficient for operational decision-making. Quality departments require not only risk forecasts but also guidance on which defect types to prioritize. CTQ analysis addresses this need by mapping dominant defect types to targeted control actions, thereby connecting daily forecasts to actionable shop-floor responses.

Unlike previous studies that report model accuracy metrics without translating forecasts into operational actions, this study proposes an end-to-end decision pipeline that connects daily defect prediction to CTQ-based quality control recommendations. First, it develops a Random-Forest-based baseline model to forecast daily defect counts for cotton yarn using routine packing-stage data under limited data conditions. Second, it demonstrates the feasibility of a data-driven risk signaling mechanism to support prioritization of daily quality control actions. Third, it integrates forecast outputs with CTQ analysis, linking daily defect risk signals to dominant defect types and actionable control recommendations through Pareto-based prioritization. Fourth, it provides an industrial case study in Indonesian textile manufacturing, offering preliminary evidence for practical data-driven quality control and a proof-of-concept workflow for comparable shop-floor environments.

The remainder of this paper is organized as follows. The next section describes the data and the proposed method, including RF modeling and evaluation metrics. The following section presents prediction results, and the CTQ analysis is shown in the next section. The discussion section outlines the operational implications of proactive quality control. Key findings and future study directions are outlined in the paper's conclusion.

2. MATERIALS AND METHODS

This study adopts a quantitative case study approach in the cotton yarn quality control department at the yarn manufacturer. The main objective is to develop a daily defect prediction model using RF regression. The model outputs are then used to support proactive quality control recommendations through a CTQ analysis. The dataset comprises historical daily production amounts and daily defect counts collected by the Quality Control (QC) and packing departments-- the data period covered from July 1 to August 31, 2025.

In this study, quality refers to the consistent ability of the production system to meet product specifications and customer expectations. Quality control is defined as a systematic set of procedures to maintain process outputs within predefined quality boundaries through measurement, evaluation, and corrective actions [4]. If quality control is ineffective, defect levels increase, and operational costs rise. Customer trust can also decline. Therefore, a proactive approach is required. Defects should not only be found after inspection. It should also forecast defect risk in advance. This enables earlier prevention and improved daily prioritization of control actions.

To address this demand, this study applies a data-driven machine learning approach. Machine learning is used to learn the relationship between input variables and a labeled output from historical data [17], [18]. The cotton yarn production at the industry consists of several connected stages, from raw fiber preparation to packaged yarn. In the blowing stage, cotton clumps are opened, and coarse impurities are removed. In the carding stage, fibers are separated and aligned to form a more uniform fiber web. In the drawing stage, multiple slivers are combined, and mass irregularity is reduced. In the open-end spinning stage, the fibers are spun into yarn and wound into cones, after which the yarn proceeds to the packing stage, where quality inspection is carried out. The input variable in this study is the daily production quantity of cotton yarn.

The target variable is the daily defect count recorded during packing inspection. This single-predictor design reflects the practical data availability constraint in the factory setting, where only production volume and defect count are recorded on a consistent daily basis [18]. While this represents a simplification of the actual causal structure underlying defect occurrence, it serves as a feasible baseline model for demonstrating the applicability of RF regression in a resource-constrained shop-floor environment [19].

The RF is used as the primary prediction algorithm. Several decision trees are used in this ensemble approach to enhance prediction stability and mitigate overfitting. Each tree is trained on a bootstrap sample of the training data. In this study, the model operates with a single input feature. The ensemble mechanism relies primarily on bootstrap sampling and output averaging across trees to stabilize predictions under limited data conditions [15]. The whole research workflow is depicted in Figure 1. The flowchart illustrates how the RF forecasts are converted into risk levels and then mapped to prioritized shop-floor actions through the CTQ matrix.

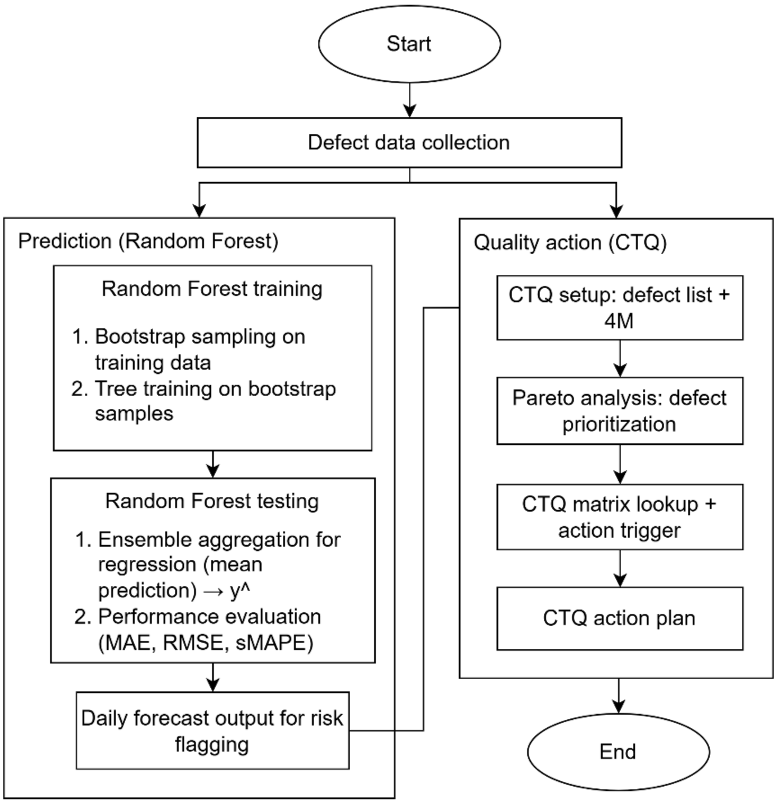

Figure 1. Overall research workflow integrating RF defect forecasting with CTQ-based action

2.1. Data Design and Preprocessing

The analyzed dataset contains two daily time series: (1) cotton yarn production volume (in cones), and (2) defect data count at the packing stage. The two data series are aligned by date. Each day has one production value and one set of defect values.

The RF model uses total daily defect count as the prediction target, aggregated across all defect types. A separate breakdown of defect counts by individual defect type was collected from the same packing inspection records and is used exclusively for Pareto analysis and CTQ matrix development. These two analyses are complementary: the RF model identifies when elevated defect risk is likely to occur, while the Pareto analysis and CTQ matrix identify which defect types to prioritize and what corrective actions to take on high-risk days.

Before modeling, the data were preprocessed to ensure reliability. First, duplicate records were removed if found. Second, data completeness was checked for each date. Observations with missing or zero values were identified and excluded to maintain a continuous daily sequence. The dataset originally covers 62 calendar days from July 1 to August 31, 2025. Two days (August 17–18, 2025) recorded zero production and zero defect values, indicating a production shutdown. These observations were excluded to maintain data integrity, resulting in 60 usable daily observations. Third, basic validity checks were performed. Negative values were not allowed. Defect counts were verified to be logically consistent with production volumes (e.g., defect count should not exceed production on the same day). These steps follow the principle that data quality has a strong influence on prediction quality.

The dataset was split using a time-ordered hold-out validation scheme. The training set consists of 50 observations (80% of data) covering July 1 to August 20, 2025, excluding the two shutdown days. The remaining 10 observations covering August 21 to August 31, 2025 were used for testing. The split was not shuffled. This choice preserves the chronological order of the time series and reduces the risk of information leakage from future days into model training. The 80/20 ratio is a common practice in applied machine learning to balance model learning and out-of-sample evaluation [16], [17].

2.2. Random Forest model

RF is an ensemble learning algorithm that constructs multiple decision trees and aggregates their outputs to produce a final prediction [12]. Each tree is trained on a bootstrap sample drawn with replacement from the training data, ensuring diversity across the ensemble. For regression tasks, the final prediction is obtained by averaging the outputs of all trees, which reduces variance and improves prediction stability compared to a single decision tree [12]. In this study, the model utilizes daily production volume as its sole input feature. The ensemble mechanism relies primarily on bootstrap sampling and output averaging across trees to stabilize predictions under limited data conditions [18].

To ensure that hyperparameters are selected through a systematic and reproducible process, an exhaustive grid search was applied. While random search offers a computationally efficient alternative by sampling randomly from the parameter space [20], grid search was preferred in this study because the parameter space is small and manageable given the limited dataset size, making full enumeration feasible [17].

To preserve the temporal structure of the data during cross-validation, a TimeSeriesSplit procedure with five folds was applied. This ensures that validation data always follows training data chronologically, preventing information leakage from future observations. The scoring criterion used was Root Mean Squared Error (RMSE). The grid search procedure is described in Algorithm 1.

Algorithm 1. Grid Search for RF Hyperparameter Tuning

Step	Description
Input	X_train: production volume (50 observations), y_train: defect count, parameter grid
Proses	
#1	Define parameter grid: n_estimators ∈ {100, 200, 300, 500}, max_depth ∈ {3, 5, 10, 15, None}, min_samples_split ∈ {2, 5, 10}, min_samples_leaf ∈ {1, 2, 4}
#2	Apply TimeSeriesSplit with 5 folds to training data
#3	For each parameter combination, train RF and evaluate mean RMSE across folds
#4	Select parameter combination with lowest mean RMSE
Output	Optimal hyperparameter configuration

The results of the grid search procedure, including the evaluation of candidate parameter combinations and the selection of the optimal configuration, are reported in the Results section. The model training and evaluation procedures are described in Algorithm 2 and Algorithm 3. Following hyperparameter selection, the RF model was trained on the training set using the optimal configuration obtained from the grid search procedure. The training process follows the standard bootstrap aggregation mechanism: each tree is trained independently on a different bootstrap sample drawn from the training data, and the final prediction is obtained by averaging outputs across all trees. This ensemble mechanism is particularly relevant in the context of limited data, as it reduces variance without requiring additional observations [12]. The training procedure is formalized in Algorithm 2.

Algorithm 2. RF model training

Step	Description
Input	X_train (50 obs), y_train, optimal hyperparameters from Algorithm 1
Process	
#1	Initialize RF with optimal hyperparameters
#2	For each of tree: draw bootstrap sample from X_train, y_train
#3	Train decision tree on bootstrap sample with optimal hyperparameters
#4	Store trained tree in ensemble
Output	Trained RF model

Algorithm 3. Model testing

Step	Description
Input	Trained model, X_train (50 obs), X_test (10 obs), y_train, y_test
Process	
#1	Generate predictions on X_train → ŷ_train
#2	Generate predictions on X_test → ŷ_test
#3	Compute MAE, RMSE, sMAPE for training set
#4	Compute MAE, RMSE, sMAPE for testing set
Output	Evaluation metrics for training and testing sets

Model performance was evaluated using three metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Symmetric Mean Absolute Percentage Error (sMAPE). MAE as Eq. (1) measures the average absolute difference between the actual and predicted values [21].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

RMSE as Eq. (2) is the square root of the mean squared error. It penalizes significant errors more heavily and is therefore sensitive to extreme deviations [22].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

sMAPE as Eq. (3) expresses relative error in percentage form. It uses a symmetric scaling with respect to the actual and predicted values [23].

$$sMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{\frac{1}{2}(|y_i| + |\hat{y}_i|)}x100\% \tag{3}$$

where,

n : Total amount of data (observations) evaluated
y_i : The observed value in the data i
$\hat{y}_i$: The value of the model's predicted value on the data i

During the training phase, the MAE, RMSE, and sMAPE values reflect how well the model learns patterns from the data used to build it. During the testing phase, these same three metrics are calculated using data not involved in the training process, thereby reflecting the model's ability to generalize to new data. Comparing metric values between the training and testing phases is also used to detect signs of overfitting. If the error values on the test data are significantly higher than those on the training data, this indicates that the model tends to memorize patterns in the training data and is less capable of generalizing to new data [24]. Conversely, relatively consistent error values between the two phases indicate that the model exhibits stable performance.

A model is said to have good predictive performance if it produces low MAE and RMSE values relative to the scale of the evaluated data, as well as an sMAPE value that falls into the low-error category according to the thresholds used in similar studies [21]. Additionally, a model's performance is considered good if it is consistent across the training and testing phases and outperforms the baseline model used in this study.

2.3. Integration with Critical to Quality (CTQ) analysis

In addition to developing a defect prediction model, this study integrates a CTQ approach. CTQ refers to the key measurable characteristics of a product or process whose performance standards or specification limits must be met in order to satisfy customer requirements [4]. The purpose of this integration is to connect model outputs with the most critical quality attributes for customers and for production performance. Within the CTQ framework, each defect type observed in cotton yarn packing is treated as a quality characteristic. Examples include crossed yarn, soiling contamination, color stains, Bad winding, slack packages, mismatched cones, trash, and yarn breaks. These defects can impact both the appearance and functionality of the product and can also disrupt production flow.

The CTQ analysis in this study consists of three steps [4]. First, each defect type is identified and defined based on packing inspection records. Second, the impact of each defect is assessed in terms of product quality and production continuity. Third, a CTQ matrix is developed. The matrix summarizes the relationship between defect types, their likely root causes, and recommended corrective or preventive actions [4]. The CTQ matrix is used as a practical decision aid that translates quality issues into specific improvement options.

The CTQ matrix is then linked to the RF prediction outputs. When the model predicts an increased defect level for a specific day, the CTQ information is used to prioritize shop-floor actions. The actions focus on the most critical defect types and their dominant causes. Examples include tighter visual inspection, improved workplace cleanliness, machine setting verification, and targeted operator guidance [4]. This linkage supports faster and more targeted responses on high-risk days. Therefore, the proposed method not only provides a statistical prediction model but also connects the prediction results to a CTQ-based action framework. This integration enables proactive quality control strategies. It helps the company focus on the quality factors that are most critical for both customers and operations.

3. RESULTS

3.1. Hyperparameter Tuning Results

The grid search procedure described in method section was applied to identify the optimal RF hyperparameter configuration. The candidate parameter space comprised four values for n_estimators {100, 200, 300, 500}, five values for max_depth {3, 5, 10, 15, None}, three values for min_samples_split {2, 5, 10}, and three values for min_samples_leaf {1, 2, 4}, resulting in a total of 4 × 5 × 3 × 3 = 180 unique combinations evaluated. Each combination was assessed using TimeSeriesSplit cross-validation with five folds on the training data, with RMSE as the scoring criterion. The candidate parameter space and the optimal configuration are presented in Table 1. The selected values represent the combination that produced the lowest mean cross-validation RMSE of 140.81.

Table 1. Final RF hyperparameter configuration

Parameter	**Candidate Values**	**Selected Value**	**Rationale**
n_estimators	100, 200, 300, 500	200	Sufficient ensemble size for stable predictions
max_depth	3, 5, 10, 15, None	3	Shallow trees prevent overfitting on small dataset
min_samples_split	2, 5, 10	2	Default minimum for node splitting
min_samples_leaf	1, 2, 4	2	Reduces variance at leaf level
bootstrap	—	True	Enables bootstrap sampling across trees
random_state	—	51	Ensures reproducibility

Best CV RMSE: 140.81

The selected max_depth of 3 is particularly noteworthy. Unrestricted tree depth risks overfitting when training data is limited, as trees may memorize noise rather than learn generalizable patterns. Constraining depth to 3 levels regularizes the model and improves out-of-sample stability, which is consistent with the proof-of-concept framing of this study [18].

3.2. Model Performance Evaluation

Following hyperparameter tuning, the RF model was trained on 50 observations and evaluated on both the training and testing sets. Model performance was assessed using three metrics: MAE, RMSE, and sMAPE. The evaluation results are presented in Table 2.

Table 2. Model evaluation results

Subset	N	MAE	RMSE	sMAPE (%)
Training	50	83.994	107.207	28.594
Testing	10	151.325	162.025	70.400

On the training set, the model achieved a MAE of 83.994, RMSE of 107.207, and sMAPE of 28.594%. These values indicate that the model was able to learn the general pattern of defect variation in the historical training data within an acceptable margin for a single-predictor baseline model.

On the testing set, the error metrics increased substantially, with an MAE of 151.325, an RMSE of 162.025, and an sMAPE of 70.400%. The considerable gap between training and testing performance suggests that the current model configuration has limited generalizability when applied to unseen data. The relatively high testing error is not unexpected given the constraints of this study, including the limited dataset consisting of only 60 daily observations, the use of a single predictor variable, and the inherent variability of daily defect counts in a shop-floor environment. Furthermore, as stated in the Methods section, this study is intended as a proof-of-concept demonstration. A more comprehensive discussion of these findings, including their practical implications and potential directions for model enhancement, is presented in the Discussion section.

3.3. CTQ Analysis and Action Framework

Prior to constructing the CTQ matrix, a Pareto analysis was conducted to identify the most critical defect types based on frequency data collected during July–August 2025. The analysis covered nine defect types recorded at the packing stage, with a total of 16,695 defect occurrences observed across the study period. The results are presented in Figure 2.

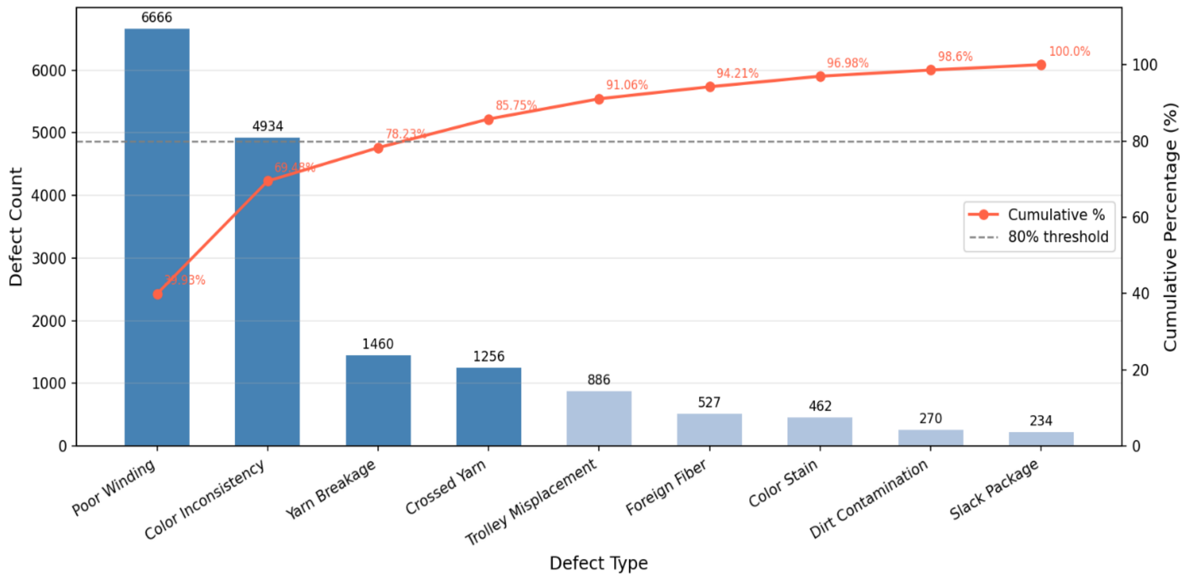

Figure 2. Pareto analysis of cotton yarn defect types (July-August 2025)

The Pareto analysis identified four defect types that collectively account for 85.75% of total defects: Bad winding (39.93%), Color Inconsistency (29.55%), Yarn Breakage (8.75%), and Crossed Yarn (7.52%). These four defect types were selected as the focus of the CTQ analysis, consistent with the Pareto principle that a small number of causes typically account for the majority of quality problems [4].

Based on the prioritized defect types, a CTQ matrix was developed. The matrix summarizes each defect type, its CTQ category, primary root cause, and recommended corrective action. The CTQ matrix is presented in Table 3.

Table 3. CTQ matrix for priority defect types

Defect Type	CTQ Category	Primary Root Cause	Recommended Action
Bad winding	Performance, Reliability	Inconsistent roller tension and unsynchronized spindle speed	Calibrate load cell and synchronize winding motor; implement 30-minute cone density inspection
Color Inconsistency	Aesthetics, Compliance	Fluctuating dyeing temperature and inconsistent fiber mixing	Calibrate thermostat and inverter; standardize mixing ratio per batch
Yarn Breakage	Performance, Reliability	Excessive thread tension and worn guide rollers	Replace worn guides; install automatic yarn breakage sensor; standardize tension at 1.0–1.2 N
Crossed Yarn	Performance, Reliability	Imbalanced roller tension and excessive winding speed	Recalibrate roller tension every 3 days; enforce RPM-to-denier SOP

The CTQ matrix serves as a practical decision aid that connects defect prioritization with targeted shop-floor actions. The integration of this matrix with the RF prediction output is discussed in the Discussion section, where the operational implications of high-risk prediction days are elaborated.

4. DISCUSSION

4.1. Model Justification and Performance Interpretation

This study employed RF regression as the primary prediction algorithm for daily defect count in cotton yarn packing. The selection of RF is justified by its ensemble mechanism, specifically bootstrap aggregation and output averaging across multiple trees, which provides variance reduction and prediction stability even under limited data conditions [12], [18]. While the single-predictor design represents a deliberate simplification driven by data availability constraints at the factory level, it reflects the practical reality of shop-floor monitoring environments where only production volume and defect count are recorded consistently on a daily basis.

The model achieved a training sMAPE of 28.594%, indicating that the ensemble was able to learn the general pattern of defect variation within the historical training data. This result confirms that the RF mechanism, even with a single input feature, partially learned the general trend of defect variation within the training period.

On the testing set, however, the sMAPE increased substantially to 70.400%, with MAE of 151.325 and RMSE of 162.025. This gap between training and testing performance warrants careful interpretation. Rather than dismissing the model as inadequate, this result should be understood within the proof-of-concept framing of this study. The testing period (August 21 to 31) comprised only 10 observations, a sample too small to draw statistically robust conclusions about generalizability. Furthermore, the defect counts observed during the testing period were notably lower and less variable than those in the training period, suggesting a distributional shift that a single-predictor model with limited training data cannot be expected to anticipate.

The testing sMAPE of 70.400% reflects a known challenge in applying RF to small datasets. For instance, studies applying RF in manufacturing quality prediction contexts note that model performance is sensitive to data volume and the representativeness of training samples [25]. However, when dataset size is limited, prediction performance is known to degrade substantially [26]. This contextualizes the testing error as a consequence of data constraints rather than a fundamental methodological failure.

These findings are consistent with Link et al. [18], who noted that ML models applied to small manufacturing datasets should be interpreted as initial demonstrations of feasibility rather than definitive evidence of production-ready predictive accuracy. The value of this study therefore lies not in the absolute accuracy of its predictions, but in demonstrating that a data-driven defect prediction framework can be operationalized using routinely available shop-floor data, providing a foundation for more comprehensive model development as richer data becomes available.

4.2 Prediction Limitations and Data Constraints

The prediction performance reported in this study is subject to several limitations that must be acknowledged transparently. The most fundamental constraint is the size of the dataset. With only 60 usable observations spanning two months, the model operates at the lower boundary of what is considered sufficient for stable machine learning estimation. Reviewer concerns regarding this limitation are well-founded: a training set of 50 observations provides limited exposure to the full range of defect variability that a production system may exhibit across seasons, shifts, operator changes, and machine conditions.

The use of a single predictor, daily production volume, further constrains the model's explanatory capacity. Defect occurrence in a packing process is plausibly influenced by a range of factors including machine condition, operator experience, shift patterns, material quality, and environmental variables such as humidity and temperature. None of these factors are captured in the current model. This is not a methodological oversight but a reflection of data availability at the time of the study. The company does not currently maintain structured digital records of these variables on a daily basis, making their inclusion infeasible without a dedicated data collection initiative.

The distributional shift observed between the training and testing periods adds another layer of complexity. Defect counts in the testing period (August 21 to 31) were consistently lower than those in the training period, which may reflect natural production variability, improved operator performance, or seasonal factors. A model trained on higher-defect periods will systematically overestimate defect counts during lower-defect periods, which is a known limitation of time-ordered hold-out validation on short time series.

These limitations collectively suggest that the current model is not yet suitable for standalone operational deployment. Its appropriate role is as a pilot framework that demonstrates the feasibility of data-driven defect monitoring and identifies the data infrastructure requirements for a more robust future implementation. Expanding the dataset to cover at least 12 months of production, incorporating additional predictor variables, and applying more sophisticated validation strategies such as rolling-origin cross-validation would substantially improve model reliability [27].

4.3 Predictive Model and CTQ Action Framework

The RF prediction output in this study is not intended to function as a standalone quality control tool. Its operational value is intended to be realized through integration with the CTQ action framework, which translates daily defect risk signals into prioritized shop-floor responses. This integration represents the core contribution of the proposed methodology.

The linkage operates as follows. When the model generates a predicted defect count for a given day, the output is interpreted as a risk signal rather than a precise numerical forecast. A predicted value that substantially exceeds the mean predicted defect count of the training period is interpreted as an elevated defect risk day, on which the CTQ matrix is activated to guide quality control priorities. The four priority defect types identified through Pareto analysis, namely Bad winding, Color Inconsistency, Yarn Breakage, and Crossed Yarn, collectively account for 85.75% of total defect occurrences. On elevated risk days, attention and preventive actions are directed toward the root causes and recommended actions associated with these defect types as outlined in Table 3.

This approach addresses a practical gap in shop-floor quality management. Traditional inspection-based quality control is reactive by nature, identifying defects only after they have occurred. By incorporating a predictive layer, the proposed framework enables quality control personnel to anticipate high-risk production days and pre-emptively intensify monitoring and corrective measures. This shift from reactive to proactive quality control is consistent with the objectives of the Six Sigma DMAIC framework, within which CTQ analysis is positioned as a tool for translating customer requirements into actionable process parameters [4].

It is acknowledged that the current threshold for elevated risk days has not been formally defined in this study, as the limited dataset size did not permit statistically robust threshold derivation. This remains an area for future development. As the prediction model is refined with larger datasets and additional predictor variables, a formal risk categorization scheme, such as low, medium, and high risk levels with corresponding CTQ response protocols, could be developed to operationalize the framework more systematically.

5. CONCLUSION

This study presents a proof-of-concept framework for proactive defect management in cotton yarn packing by integrating a RF regression model with a CTQ-based action framework. The model was trained on 50 daily observations using a single predictor, daily production volume, with hyperparameters systematically selected through grid search and TimeSeriesSplit cross-validation. Training performance indicated partial in-sample learning with a sMAPE of 28.594%, while testing sMAPE of 70.400% reflects the constraints of limited data and a distributional shift between periods.

Pareto analysis identified four priority defect types: Bad Winding, Color Inconsistency, Yarn Breakage, and Crossed Yarn, collectively accounting for 85.75% of total defect occurrences. These were incorporated into a CTQ matrix linking each defect type to its primary root causes and recommended corrective actions. The integration of RF prediction outputs with the CTQ matrix provides a structured mechanism for directing shop-floor responses on elevated defect risk days, supporting a shift from reactive to proactive quality control.

Future work should prioritize expanding the dataset to cover longer production periods and incorporating additional predictor variables. Exploring alternative predictive methods better suited for small datasets would provide a comparative basis for evaluating model in situations with data-limited conditions.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Universitas Setia Budi for the financial support provided for this research and the publication of this article.

REFERENCES

[1] Kemenperin, "Kemenperin bidik ekspor industri TPT capai USD14 miliar," Kementerian Perindustrian, 2022. [Online]. Available: https://www.kemenperin.go.id/artikel/23452/Kemenperin-Bidik-Ekspor-Industri-TPT-Capai-USD-14-Miliar

[2] A. Wirajaya, "Kondisi Industri Tekstil Makin Terguncang, Ini Solusinya," 2025. [Online]. Available: https://enciety.co/kondisi-industri-tekstil-makin-terguncang-ini-solusinya/

[3] T. Pyzdek and P. A. Keller, *The six sigma handbook*, 5th Edition. New York: McGraw-Hill Education, 2018. [Online]. Available: https://www.accessengineeringlibrary.com/content/book/9781260121827

[4] D. C. Montgomery, *Introduction to Statistical Quality Control*, 8th edition. John Wiley & Sons, 2020. [Online]. Available: https://www.wiley.com/en-us/Introduction+to+Statistical+Quality+Control,+8th+Edition-p-9781119723097

[5] D. Harianto, H. Pujianto, H. S. Bintang, and D. Alfanti, "Analisa Perbaikan Benang Kusut Akibat Ring Touch Pada Proses Penggulungan Di Mesin Ring Spinning Frame," *JISI: Jurnal Integrasi Sistem Industri*, vol. 12, no. 1, pp. 67–78, 2025, doi: 10.24853/jisi.12.1.67-78.

[6] N. K. Sahoo, "Efficiency Improvement by Reducing Rework and Rejection on the Shop Floor," *International Journal of Engineering Research and*, vol. V9, no. 06, pp. 1185–1191, 2020, doi: 10.17577/ijertv9is060857.

[7] Rohmat, Sudarmiatin, and T. S. Dhewi, "The Influence of Product Quality on Consumer Loyalty with Customer Satisfaction and Brand Trust as Mediating Variables (Case Research on Samsung Smartphone Consumers in Malang City)," *International Journal Of Humanities Education and Social Sciences (IJHESS)*, vol. 2, no. 1, pp. 225–238, 2022, doi: 10.55227/ijhess.v2i1.232.

[8] W. Wahyudin, D. Herwanto, and N. Helmy, "Perbaikan Kualitas Pembuatan Benang Cotton Ne 40 S Dengan Menggunakan Metode Six Sigma-DPOM Di PT. XYZ," *Barometer*, vol. 2, no. 2, pp. 57–59, 2017, doi: 10.35261/barometer.v2i2.906.

[9] A. Perwitasari, A. T. Wahyudi, and W. Widhiarso, "Strategi Pengendalian Produk Cacat di Industri Penyamakan Kulit melalui Pendekatan DMAIC Six Sigma," in *Prosiding Sains dan Teknologi*, Kabupaten Bekasi, Jawa Barat: Fakultas Teknik, Universitas Pelita Bangsa, Feb. 2025, pp. 612–623. Accessed: Dec. 12, 2025. [Online]. Available: https://ejurnal.pelitabangsa.ac.id/index.php/SAINTEK/article/view/5609

[10] S. Iskandar, V. G. V. Putra, and H. Hermansyah, "Prediksi End Breakage Benang Kapas Di Mesin Rotor Spinning Menggunakan Pendekatan Jaringan Saraf Tiruan," *EduFisika: Jurnal Pendidikan Fisika*, vol. 7, no. 1, pp. 72–87, 2022, doi: 10.59052/edufisika.v7i1.19543.

[11] H. K. Choi, W. Lee, S. M. M. Sajadieh, S. D. Noh, H. S. Son, and S. B. Sim, "Machine Learning Algorithms for Process Optimization and Quality Prediction of Spinning in Textile Industries," in *Product Lifecycle Management. Leveraging Digital Twins, Circular Economy, and Knowledge Management for Sustainable Innovation*, vol. 702, in IFIP Advances in Information and Communication Technology, vol. 702. , Cham: Springer Nature Switzerland, 2024, pp. 221–232. doi: 10.1007/978-3-031-62582-4_20.

[12] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[13] M. B. Kursa, "Robustness of Random Forest-based Gene Selection Methods," *BMC Bioinformatics*, vol. 15, no. 1, 2014, doi: 10.1186/1471-2105-15-8.

[14] C. Junpradub and K. Asawarungsaengkul, "Predicting the Product Classification of Hot Rolled Steel Sheets Using Machine Learning Algorithms," *Engineering Journal*, vol. 27, no. 8, pp. 51–61, Aug. 2023, doi: 10.4186/ej.2023.27.8.51.

[15] M. A. Al-Ebrahim and R. S. Ransing, "A Scalable Random Forest (SRF) Approach for Non-Linear Predictive Modelling using Small Manufacturing Datasets," *Journal of Intelligent Manufacturing*, Jun. 2026, doi: 10.1007/s10845-026-02834-2.

[16] N. C. Igbokwe, C. O. Nwamekwe, and P. S. Aguh, "Predictive Modeling of Manufacturing Defects using Machine Learning: A Comparative Performance Study in a Manufacturing SME," *African Journal of Advances in Engineering and Technology*, vol. 1, no. 2, pp. 93–115, 2025, doi: https://doi.org/10.31920/2978-3224/2025/V1N2A5.

[17] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011, doi: http://jmlr.org/papers/v12/pedregosa11a.html.

[18] P. Link *et al.*, "Capturing and incorporating expert knowledge into machine learning models for quality prediction in manufacturing," *Journal of Intelligent Manufacturing*, vol. 33, no. 7, pp. 2129–2142, Oct. 2022, doi: 10.1007/s10845-022-01975-4.

[19] Y. Yang and H. Wang, "Random Forest-Based Machine Failure Prediction: A Performance Comparison," *Applied Sciences*, vol. 15, no. 16, p. 8841, Aug. 2025, doi: 10.3390/app15168841.

[20] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," *Journal of Machine Learning Research*, vol. 13, no. 10, pp. 281–305, 2012, doi: http://jmlr.org/papers/v13/bergstra12a.html.

[21] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005, doi: https://doi.org/10.3354/cr030079.

[22] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014, doi: 10.5194/gmd-7-1247-2014.

[23] S. Makridakis, "Accuracy measures: theoretical and practical concerns," *International Journal of Forecasting*, vol. 9, pp. 527–529, 1993, doi: https://doi.org/10.1016/0169-2070(93)90079-3.

[24] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, pp. 679–688, 2006, doi: 10.1016/j.ijforecast.2006.03.001.

[25] M. Schroth *et al.*, "Optimization paper production through digitalization by developing an assistance system for machine operators including quality forecast: a concept." arXiv, 2022. doi: 10.48550/ARXIV.2206.11581.

[26] S. Han, B. D. Williamson, and Y. Fong, "Improving Random Forest Predictions in Small Datasets from Two-Phase Sampling Designs," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 322, Dec. 2021, doi: 10.1186/s12911-021-01688-3.

[27] L. J. Tashman, "Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review," *International Journal of Forecasting*, vol. 16, no. 4, pp. 437–450, Oct. 2000, doi: 10.1016/S0169-2070(00)00065-0.

DECLARATION OF AI-TOOL USAGE

The authors used AI-based tools solely to assist with language editing, including language refinement, and grammar checking. All key intellectual contributions, including research design, data analysis, and interpretation of findings, were conducted independently by the authors. The authors take full responsibility for the content of this manuscript.