Engineering and Technology Journal e-ISSN: 2456-3358

Volume 11 Issue 06 June-2026, Page No.- 10488- 10495

DOI: 10.47191/etj/v11i06.24, I.F. – 8.595

External Validation of Ensemble Learning Models on Unseen Data for Production Output Prediction: A Comparative Study Against Garment Industry Theoretical Targets

Hari Murti[1], Heni Candra Kirana[2], Widiyanto Tri Handoko[3], Eko Nur Wahyudi[4], Eka Ardhianto[5]

[1] Program Study of Information Systems, Faculty of Information Technology and Industry, Universitas Stikubank, Semarang, Central Java, Indonesia

[2,5]Program Study of Information Technology, Faculty of Information Technology and Industry, Universitas Stikubank, Semarang, Central Java, Indonesia

[3,4]Program Study of Information Engineering, Faculty of Information Technology and Industry, Universitas Stikubank, Semarang, Central Java, Indonesia

ABSTRACT: The garment industry requires accurate production output predictions to minimize operational inefficiencies and support adaptive production planning. However, static, manual production targets often fail to reflect the real-time dynamics of the production floor, typically resulting in overestimations. This study evaluates the accuracy, stability, and generalization capability of ensemble learning models via external validation on unseen operational data, while comparing their performance against traditional theoretical targets. Employing a quantitative experimental approach, the study utilizes 700 historical data points for model training and 60 recent observations as an external test dataset, incorporating manpower and the Standard Minute Value (SMV) gap as key variables. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). The results demonstrate that both models exhibit robust performance, achieving R^2 values above 0.90 on unseen data. Comparatively, Random Forest achieved an MAE of 6.008, RMSE of 8.56, MAPE of 23.29%, and an R^2 of 0.908, whereas Gradient Boosting yielded an MAE of 5.892, RMSE of 8.05, MAPE of 25.60%, and an R^2 of 0.919. Although Gradient Boosting outperformed Random Forest in absolute error metrics and R^2, Random Forest demonstrated superior relative prediction stability, winning 51.67% of daily prediction comparisons in a "Battle of Models" framework. Visual analysis using scatter plots and box plots confirmed that Random Forest maintains a more consistent error distribution and captures actual output fluctuations more realistically than the company's manual targets. These findings indicate that integrating the SMV gap variable effectively captures hidden losses within the production process. Consequently, Random Forest is recommended as the foundation for developing a data-driven Decision Support System (DSS) to facilitate more adaptive, realistic, and efficient production target setting in the garment industry.

KEYWORDS: Ensemble Learning, Machine Learning, Gradient Boosting, Random Forest, External Validation

I. INTRODUCTION

The garment manufacturing industry is characterized by high operational dynamics driven by the transition toward Industry 4.0. As a labor-intensive sector, accurate production output prediction is a crucial aspect of maintaining cost efficiency and ensuring timely delivery [1], [2]. The implementation of Industry 4.0 technologies fosters the integration of the Internet of Things (IoT) and machine learning within manufacturing systems to support automation, real-time data processing, and enhanced operational production efficiency [3]. In line with this digital transformation, machine learning applications are increasingly utilized to optimize operational efficiency [4], [5]. This aligns with recent research indicating that integrating machine learning into smart manufacturing enhances process efficiency and data-driven decision-making capabilities [6]. Conversely, establishing static production targets without considering real-time conditions often triggers reactive overtime policies, which inflate operational costs without a proportional increase in productivity [7].

Conventionally, achieving production targets is often assumed to depend heavily on manpower expansion. However, literature indicates that the impact of labor on output is not always linear, as the stability of operator performance frequently exerts a greater influence than simply increasing headcount. Interestingly, a previous study [8] demonstrated contrasting results; through historical data analysis using Ensemble Learning methods, it was found that the actual Standard Minute Value (SMV) and the SMV gap exerted a more dominant influence on production output than

manpower. This finding underscores that task complexity and operational time efficiency are more critical determinants of production attainment. This condition is supported by prior studies showing that SMV [9], [10] and manpower fluctuations [11], [12] serve as primary factors behind failed production targets. This premise is further reinforced by a recent study demonstrating that process time deviation is inversely proportional to daily output and correlates significantly with productivity declines [12].

The application of artificial intelligence in intelligent manufacturing can optimize managerial efficiency through real-time data analysis and adaptive decision-making, while simultaneously mitigating the inefficiencies of conventional methods [3]. On the other hand, the limitations of traditional regression models in capturing non-linear patterns within operational data have driven the adoption of ensemble tree algorithms. Models such as Random Forest and Gradient Boosting have proven to be superior and highly adaptive to data complexity and noise [13], [14], [15].

Although Random Forest and Gradient Boosting exhibit robust performance during internal testing, their effectiveness in confronting actual production dynamics remains to be validated. One of the primary challenges of implementing machine learning in industrial settings is the risk of overfitting, wherein a model performs well on training data but falls short on new, unseen data. Most current research still relies heavily on internal validation, leaving models vulnerable to performance degradation due to overfitting [16]. Consequently, an evaluation based on the latest operational conditions is critical before these algorithms can be practically adopted by management.

To bridge this research gap, this study focuses on stability testing and the external validation of models using recent operational datasets (unseen data). The evaluation is extended to the implementation phase under actual production conditions to comprehensively assess algorithmic reliability. The novelty of this research lies in the direct comparison between model predictions and the company's conventional theoretical estimations. This approach enables a robust assessment of the model's reliability in mitigating the risks of failing to meet production targets.

Based on this background, this study aims to address two primary research questions:

1. How do the Random Forest and Gradient Boosting models perform in terms of accuracy and stability during external validation using the latest production data (unseen data)?
2. How effective are Ensemble Learning-based predictions when compared to the company's theoretical targets?

Ultimately, this study aims to identify the most optimal predictive algorithm under real-world conditions. The practical implications of this study's outcomes are expected to serve as a strategic foundation for designing a more precise and adaptive Decision Support System (DSS) for garment industry management.

II. METHOD

This study employs a computational experimental approach to evaluate the stability of predictive models under real-world operational conditions. Distinct from prior research that focused heavily on model development and feature importance analysis to identify dominant factors [8], the present study specifically targets external validation to assess the model's generalization capability on entirely unseen datasets.

The Random Forest and Gradient Boosting algorithms, which underwent training and optimization in a preceding study, are reused in this research. In general, the research workflow consists of two primary stages: the model development stage and the external validation stage, as illustrated in Figure 1.

Figure 1. Framework of the Model's External Validation Workflow.

Referring to Figure 1, the overall research framework is divided into two continuous phases. In Phase 1 (prior research), the models were trained using historical datasets through optimization and hyperparameter tuning up to the initial evaluation stage. Subsequently, the focus of the present study centers on Phase 2, which entails the external validation scheme. Broadly speaking, this second phase encompasses: (1) the utilization of new, unseen data, (2) the output prediction process using the pre-trained models without retraining (no retraining), (3) the evaluation of regression performance using MAE, MAPE, RMSE, and R^2 metrics, and (4) a comprehensive comparison between the machine learning predictions, theoretical targets, and actual outputs. This entire sequence is designed to comprehensively evaluate model reliability when deployed under actual operational conditions.

A. Dataset and Feature Engineering

To support the external validation scheme, the research dataset is classified into two primary groups: 700 historical observations utilized as the foundation for model training and optimization in the prior study [8], and 60 recent observations serving as the external test dataset (unseen data). These recent data points are extracted from a separate operational period, thereby accurately representing actual and dynamic production conditions. Employing a time-split validation approach for the test data aims to evaluate the models'

capability to generalize to dynamic production patterns that were not encountered during the training phase.

All observations were recorded at a 20-minute time interval resolution. The selection of this interval is based on the standard production cycle at the case study company, while concurrently serving as an effective instrument for daily production monitoring. Maintaining a consistent 20-minute interval aligns with the methodology established in the predecessor study [8], thereby ensuring continuity in the production data analysis.

The variables employed in this study consist of independent and dependent variables. The independent variables include manpower, which represents the workforce size, and the Standard Minute Value (SMV) gap, which is defined as the difference between the theoretical operation breakdown SMV and the actual shop-floor SMV. Meanwhile, the dependent variable is the daily production output measured in units (pieces/pcs). Implementing feature engineering on the SMV gap is considered critical, as it accurately captures operational efficiency deviations and mitigates the impact of data noise, thereby enhancing the quality of information leveraged by the ensemble learning models to yield more precise predictions.

B. Models and Prediction Procedures

The model architecture utilized in this study encompasses two ensemble learning algorithms, namely Random Forest and Gradient Boosting. Both models are derived from prior research and have undergone hyperparameter optimization using the Grid Search technique alongside k-fold cross-validation [8].

During the external validation phase, the models are deployed without retraining to test performance consistency and generalization capabilities on the latest operational data (unseen data) in inference mode. The workflow of the model prediction process is presented in Figure 2.

Figure 2. Workflow of the Model Prediction Process.

Based on Figure 2, the process initiates by feeding the latest operational data into the pre-trained Random Forest and Gradient Boosting models. The models then generate production output predictions, which are subsequently analyzed through two evaluation approaches. The first approach compares the predicted results against the actual production output to calculate the prediction error rate using regression metrics, including MAE, MAPE, RMSE, and R^2. The second approach conducts a direct head-to-head comparison between the model predictions and the company's theoretical targets to evaluate the model's relevance within the context of industrial operations.

These theoretical targets are derived from conventional manual estimation methods historically used in the production planning process. This method serves as a baseline to determine whether the machine learning-based approach can produce estimates that are more adaptive and closer to real-world conditions compared to the manual approach, which tends to be static.

C. Performance Evaluation Metrics

Model performance evaluation is conducted using four regression statistics to provide a comprehensive assessment of prediction accuracy and reliability. Utilizing a combination of these metrics aims to examine prediction errors from the perspectives of absolute values, percentages, and sensitivity to extreme deviations.

1. Mean Absolute Error (MAE)

MAE represents the average of the absolute differences between the predicted and actual values. This metric measures the prediction error level based on the absolute mean difference between actual data and predicted data [17]. In garment operations, MAE provides a direct representation of the average output deviation in units (pcs). Consequently, a lower MAE value indicates a superior level of model prediction accuracy.

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

2. Mean Absolute Percentage Error (MAPE)

MAPE measures the average percentage of prediction errors relative to the actual values. This metric is highly crucial for management to evaluate model performance relative to production volume.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2)$$

3. Root Mean Square Error (RMSE)

RMSE is used to measure the error magnitude between the actual values and model predictions. By squaring the errors, this metric penalizes larger errors more heavily, making it highly sensitive to outliers. A lower RMSE value indicates enhanced prediction accuracy [18]. This metric is highly relevant for evaluating model stability under fluctuating production conditions, such as machine breakdowns or mass operator absenteeism.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (3)$$

4. R-Squared (R^2)

R-Squared R^2) is a measure that reflects the relationship between actual and predicted values within a model [17]. This metric is utilized to quantify the extent to which independent variables, such as manpower and the SMV gap, can explain

the variance in production output. An R^2 value approaching 1 indicates that the model possesses superior predictive and generalization capabilities regarding production data patterns.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (4)$$

Where:

n : Total number of observations (60 data points)

y_i : Actual production output value at the i-th observation

$\hat{y}_i$: Model-predicted output value

$\bar{y}_i$: Mean production output value of all actual data

The combination of these four metrics allows for a more holistic evaluation. The optimal model is determined based on the lowest error values (MAE, MAPE, RMSE) alongside the most stable R^2 coefficient. Through this approach, the research can identify the most reliable algorithm to support data-driven decision-making within the garment industry environment.

D. External Validation Scheme

External validation in this study utilizes an independent dataset completely isolated from both the training and internal testing phases. Executing external validation is essential to measure the model's generalization capacity on unseen data and to evaluate model robustness when faced with shifts in data distribution that could trigger a decline in prediction performance [19].

Beyond mitigating the risk of overfitting, external validation is also utilized to identify potential model decay resulting from temporal shifts in data characteristics [20]. This process is vital for assessing model reliability under dynamic, real-world manufacturing conditions, such as operator performance fluctuations and technical disruptions. Consequently, the validation outcomes provide not only a statistical evaluation but also demonstrate the model's readiness for deployment as a Decision Support System (DSS) to facilitate more precise and adaptive production planning in the garment industry.

III. RESULTS AND DISCUSSIONS

A. External Validation on Recent Production Data (Unseen Data)

This stage evaluates model reliability using 60 observations of the latest operational data to validate the consistency of the manpower and SMV gap features identified in a previous study [8]. This external validation aims to ensure the model's generalizability to data that were not involved in the training process, thereby preventing overfitting.

The utilization of chronologically collected data reflects actual production dynamics. These evaluation results serve as the primary benchmark for determining the algorithm with the highest stability and precision to support data-driven production planning. To maintain the integrity of the results, data preprocessing procedures strictly adhere to the cleaning and variable handling protocols validated previously [8]. This consistent data pipeline ensures that the evaluation outcomes accurately reflect the true performance of the algorithm when processing recent operational data, without introducing processing bias.

B. Statistical Performance Analysis of the Models

The statistical performance of the Random Forest and Gradient Boosting algorithms was evaluated using 60 observations of actual data across four primary metrics, as presented in Table 1.

Table 1. Summary of Model Validation Results on 60 New Data Observations.

Evaluation Metrics	Random Forest	Gradient Boosting
MAE (Mean Absolute Error)	6.008	5.892
RMSE (Root Mean Squared Error)	8.56	8.05
MAPE (Mean Absolute Percentage Error)	23.29%	25.60%
R-Squared (R^2)	0.908	0.919

Based on Table 1, both models exhibit robust performance, with R^2 values exceeding 0.90. This confirms that the manpower and SMV gap variables contribute significantly, explaining more than 90% of the variance in production output.

Comparatively, Gradient Boosting outperforms Random Forest across the majority of absolute accuracy metrics (RMSE, MAE, and R^2). This demonstrates the effectiveness of the sequential learning approach in Gradient Boosting for precisely capturing non-linear relationships and minimizing extreme prediction errors compared to the parallel approach utilized in Random Forest.

Nevertheless, Random Forest achieves superior performance on the MAPE metric, registering the lowest value at 23.29%. This indicates that Random Forest exhibits greater stability in maintaining relative errors against fluctuations in production scale. Given that the differences in statistical performance between the two models are marginal, further visual analysis is required to thoroughly evaluate the consistency of the error distribution.

C. Comparison of Predictions against Actual Output and Targets

The evaluation phase continues with a visual analysis to assess the models' capability to track production fluctuations, thereby complementing the preceding statistical evaluation results. Figure 3 illustrates the comparison among the actual output, the predictions generated by the ensemble learning models (Random Forest and Gradient Boosting), and the company's manual targets.

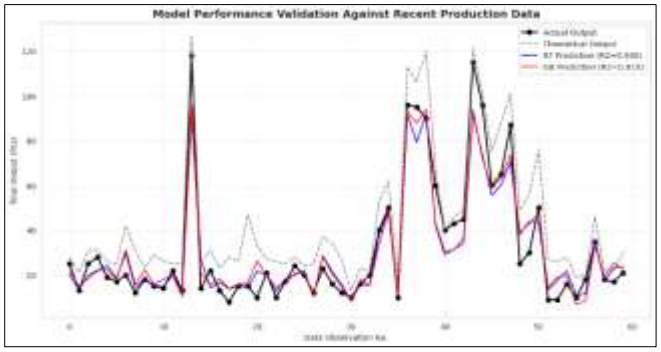

Figure 3. Visualization of Model Prediction Comparison against Actual Output and Manual Targets.

Based on the visualization in Figure 3, both machine learning models demonstrate high adaptability in tracking the dynamics of output fluctuations on the production floor. The prediction lines for Random Forest and Gradient Boosting closely align with the actual output line across all 60 observation points. This consistency demonstrates that integrating the manpower and SMV gap variables provides robust predictors for precisely determining daily production realization.

A significant contrast is observed in the manual target line (gray dashed line), which tends to remain static and frequently sets overly optimistic (overestimated) targets. These conventional targets appear to overlook actual floor-level constraints, thereby creating a wide discrepancy relative to production reality. Conversely, the ensemble learning approach offers more realistic predictions, minimizing the risk of setting targets that fail to accommodate resource constraints. Visually, Random Forest exhibits a smoother response in capturing sharp data fluctuations, reinforcing the previous findings regarding the MAPE metric.

D. Prediction Distribution Analysis (Scatter Plot)

To strengthen spatial validation, a prediction distribution analysis was conducted using a scatter plot, as illustrated in Figure 4. This graph maps the relationship between the actual output values on the X-axis and the predicted values on the Y-axis, where the diagonal line (Y = X) represents perfect precision.

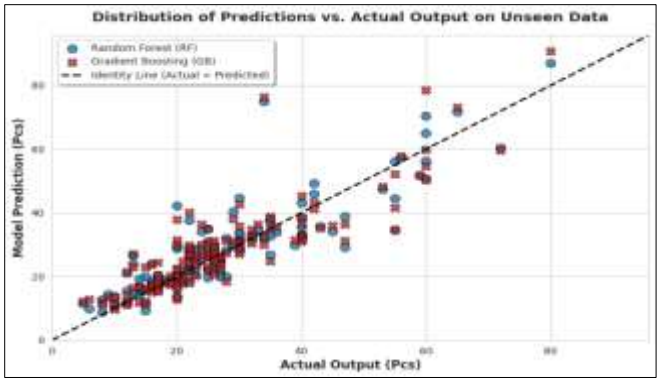

Figure 4. Scatter Plot Comparison between Actual Output and Predictions of Random Forest and Gradient Boosting Models.

The visualization in Figure 4 demonstrates that the prediction points for both algorithms are densely clustered around the ideal line (Y = X). This distribution pattern confirms the high accuracy previously indicated by R^2 values exceeding 0.90, proving that the models exhibit minimal error across all production volume tiers.

Comparatively, the Random Forest model displays a more consistent density in the medium to high output range than Gradient Boosting. The absence of extreme outliers deviating from the diagonal line indicates that the models have successfully captured the underlying data patterns. Utilizing Random Forest as the selected model offers the additional advantage of lowering the risk of critical errors in daily operations, making it a reliable choice for implementing a decision support system on the production floor.

E. Stability and Error Distribution Analysis

To reinforce the statistical evaluation results, a stability analysis was conducted on the absolute error distribution using the box plot illustrated in Figure 5. This analysis aims to assess model consistency in maintaining prediction error levels within production management tolerance limits.

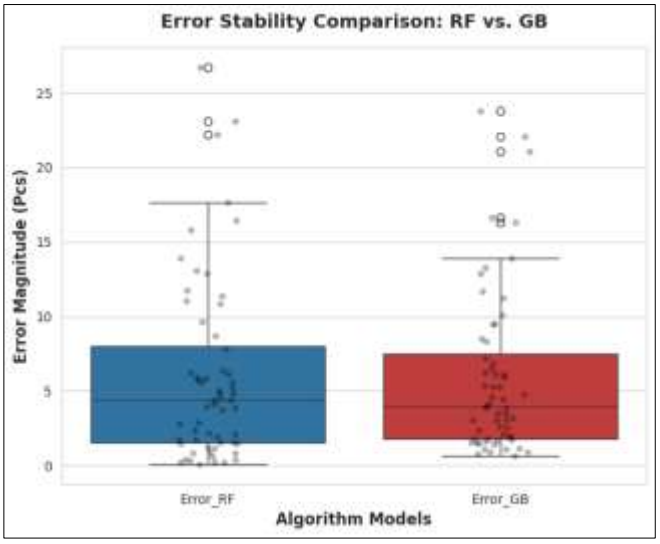

Figure 5. Comparison of Error Stability between Random Forest and Gradient Boosting Algorithms.

Based on the visualization in Figure 5, the error distributions for both ensemble learning models demonstrate competitive stability. The median error for both algorithms remains within a low range, specifically below 5 units per 20-minute production interval. This indicates that the majority of the model estimates achieve high precision relative to actual floor-level conditions.

Although both models exhibit similar overall performance, Random Forest shows a slightly more consistent error distribution range within the lower quartile region. The presence of several outliers with error values exceeding 15 units in both models represents unexpected operational anomalies on the production floor, such as machine

breakdowns or other non-technical factors. However, statistically, the clustering of most error data at low values demonstrates that the proposed machine learning approach possesses excellent robustness in handling production data dynamics without experiencing extreme prediction error fluctuations.

F. Model Generalization Analysis

This stage evaluates the generalization capability of the ensemble learning models to maintain predictive performance on new data (unseen data). The evaluation is conducted by comparing the baseline performance of the development data from the previous study against the external validation data in this study. This performance comparison is presented in Table 2 and visualized in Figure 6.

Table 2. Model Performance Comparison between Development and External Validation Data.

Evaluation Metrics	Random Forest	Gradient Boosting
MAE (Mean Absolute Error)	4.55 → 6.008 (+1.458)	4.88 → 5.892 (+1.012)
RMSE (Root Mean Squared Error)	6.85 → 8.56 (+1.71)	7.21 → 8.05 (+0.84)
MAPE (Mea Absolute Percentag Error)	19.12% → 23.29% (+4.17%)	20.78% →25.60% (+4.82%)
R-Squared (R^2)	0.758 → 0.908 (+0.150)	0.733 → 0.919 (+0.186)

Note: Values are displayed in the format: [Development Data] → [Validation Data] (Delta).

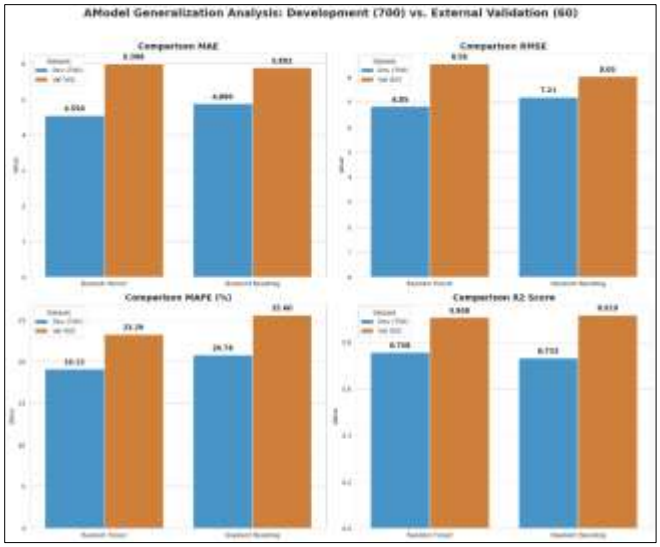

Figure 6. Model Performance Comparison between Development Data and External Validation Data.

Based on Table 2 and Figure 6, both algorithms generally exhibit an increase in error metrics (MAE, RMSE, and MAPE) when tested on external validation data. For instance, the MAPE for the Random Forest model increased by 4.17% (from 19.12% to 23.29%), while that of Gradient Boosting increased by 4.82%. This trend indicates a standard performance decay resulting from shifting operational data dynamics relative to the training dataset. This moderate performance decline demonstrates that neither model suffers from extreme overfitting.

Interestingly, despite the rise in absolute errors, the R^2 values for both models increased significantly, exceeding 0.90 (as shown in the bottom-right quadrant of Figure 6). This positive anomaly confirms that the manpower and SMV gap features possess a substantially stronger and more relevant explanatory capacity for mapping production variance patterns under current conditions. Consequently, the models remain robust in capturing primary production trends even when encountering more complex unseen data, confirming excellent generalization capability for deployment in real-world operations.

G. Model Superiority Analysis (Battle of Models)

This stage evaluates point-to-point predictive performance through a direct comparison mechanism designated as the "Battle of Models." This evaluation compares the estimations of the Random Forest and Gradient Boosting algorithms head-to-head across 60 recent data observation points to identify the model achieving the highest frequency of accuracy relative to actual daily values.

Table 3. Summary of Battle of Models Results (60 Observations)

Performance Indicator	Random Forest	Gradient Boosting
Frequency of Most Accurate Prediction	31 Observasi	29 Observasi
Win Percentage	51.67%	48.33%

Based on Table 3, Random Forest demonstrates frequency-based dominance, winning 31 out of 60 observations, which translates to 51.67 Although Gradient Boosting remains highly competitive, these results indicate that Random Forest is more consistent in delivering estimations closest to the actual values within daily operational frequencies.

H. Discussion

The determination of the optimal model in this study was conducted by aligning all evaluation results, ranging from statistical accuracy to visual data distribution analysis. Although Gradient Boosting exhibits slightly higher precision in absolute error metrics, Random Forest proves superior in maintaining relative error stability and winning the frequency of accuracy through the "Battle of Models" mechanism. Support from the scatter plot and box plot analyses further solidifies Random Forest's position due to its higher prediction consistency and more stable error range. This superiority is facilitated by the algorithm's architecture, which effectively mitigates disruptions on the production

floor, thereby remaining robust against dynamic data variations without experiencing severe performance decay [14].

The models' capability to track actual output fluctuations indicates that the company's manual targets have historically been unrealistic. Conventional targets tend to be overly optimistic because they are based solely on machine capacity without considering actual field-level constraints. Conversely, the machine learning approach captures hidden operational factors through the SMV gap variable, which represents operator fatigue and technical bottlenecks during the production process. Given its strong generalization capability on recent data, this model generates more realistic predictions that accurately reflect true production capacity.

From a managerial perspective, these findings provide a foundation for the company to transition from static target setting to data-driven production planning. The discrepancy between model predictions and theoretical targets indicates that conventional targets must be adjusted to actual operational conditions through a more adaptive decision support system. Implementing dynamic targets can enhance production planning accuracy, reduce unplanned overtime, and help balance organizational efficiency with worker workload on the production floor.

IV. CONCLUSIONS

Random Forest is the optimal model for predicting garment production output, achieving an R^2 value exceeding 0.90 during external validation. This demonstrates superior consistency in prediction accuracy and stability compared to Gradient Boosting when processing dynamic operational data. Furthermore, the ensemble learning approach proves more effective and realistic than the company's conventional theoretical targets, which tend to be static and overly optimistic. This efficacy stems from the model's capacity to integrate the SMV gap variable, thereby capturing hidden losses that remain unidentified by conventional methods. Overall, the model's ability to represent actual production capacity establishes a strategic foundation for adopting a dynamic target system driven by a Decision Support System (DSS). This implementation enhances production planning precision while balancing industrial efficiency with realistic operator workloads.

REFERENCES

1. T. Martina, I. Fauzi, U. Pramesvari, A. Ramadhan, and L. N. Asri, "Perancangan Order Management System Berbasis Web Application untuk IKM Garmen," *Journal of Community Services in Sustainability*, vol. 2, no. 1, pp. 1–10, 2024, doi: 10.52330/jocss.v2i1.178.
2. I. T. Kartika and N. Azizah, "Peran Industri Garmen sebagai Motor Pemberdayaan Perempuan di Bangladesh: Analisis Indikator," *Jurnal Hubungan Internasional*, vol. 2, no. 1, pp. 261–281, Mar. 2025, doi: 10.36722/mondial.v2i1.3746.
3. G. Gallitto *et al.*, "External Validation of Machine Learning Models - Registered Models and Adaptive Sample Splitting," Dec. 2023, doi: 10.1101/2023.12.01.569626.
4. H. Tercan and T. Meisen, "Machine Learning and Deep Learning Based Predictive Quality in Manufacturing: A Systematic Review," *J. Intell. Manuf.*, vol. 33, pp. 1879–1905, May 2022, doi: 10.1007/s10845-022-01963-8.
5. I. Ayu, A. Fudoli, and M. H. Fahamsyah, "Metode Demand Forecasting dalam Menjalankan Manajemen Operasi pada Industri Manufaktur," *EKOMABIS: Jurnal Ekonomi Manajemen Bisnis*, vol. 3, no. 2, pp. 127–136, Aug. 2023, doi: 10.37366/ekomabis.v3i02.286.
6. B. Ördek, Y. Borgianni, and E. Coatanea, "Machine Learning-Supported Manufacturing: A Review and Directions for Future Research," *Prod. Manuf. Res.*, vol. 12, no. 1, p. 1, 2024, doi: 10.1080/21693277.2024.2326526.
7. A. Anugrah, H. H. Mohamad, J. Otniel, M. R. Fahrezi, M. Radian, and F. Siswajanthy, "Analisis Industri Tekstil Di Jawa Barat Sebelum Dan Setelah Krisis Ekonomi," *Doktrin: Jurnal Dunia Ilmu Hukum dan Politik*, vol. 2, no. 2, pp. 118–135, Apr. 2024, doi: 10.59581/Doktrin-widyakarya.v2i1.2579.
8. H. C. Kirana and E. Ardhianto, "Feature Importance Analysis of SMV Gap and Manpower Variables on Garment Production Output based on Ensemble Learning," *Sistemasi: Jurnal Sistem Informasi*, vol. 15, no. 5, pp. 1796–1806, May 2026, doi: 10.32520/stmsi.v15i5.
9. Kamruzzaman, T. A. T. Tutul, I. Hasan, A.-A. Prodhan, K. A. Tahmid, and A. Taher, "An Analysis of Standard Minute Value (SMV) to Increase Productivity in the Sewing Division: A Case Study on Jeans Pant Production," *International Conference on Mechanical, Industrial and Materials Engineering*, pp. 1–6, Dec. 2024.
10. B. Bizuneh and R. Omer, "Lean Waste Prioritisation and Reduction in the Apparel Industry: Application of Waste Assessment Model and Value Stream Mapping," *Cogent Eng.*, vol. 11, no. 1, pp. 1–22, Apr. 2024, doi: 10.1080/23311916.2024.2341538.
11. M. Ewnetu and Y. Gzate, "Assembly Operation Productivity Improvement for Garment Production Industry Through the Integration of Lean and Work-Study, a Case Study on Bahir Dar Textile Share Company in Garment, Bahir Dar, Ethiopia," *Heliyon*, vol. 9, pp. 1–13, Jul. 2023, doi: 10.1016/j.heliyon.2023.e17917.

12. P. K. Dewa, R. Meilina, S. Budiman, and I. N. Afiah, "Optimasi Capaian Target Produksi Melalui Peningkatan Faktor Manusia," *Jurnal Penelitian dan Aplikasi Sistem dan Teknik Industri*, vol. 18, no. 1, pp. 43–52, Apr. 2024, doi: 10.22441/pasti.2024.v18i1.005.
13. Y. A. Purmala, "Penerapan machine learning dalam meningkatkan produktivitas di industri manufaktur: Tinjauan literatur," *Operations Excellence: Journal of Applied Industrial Engineering*, vol. 13, no. 2, pp. 267–275, Jul. 2021.
14. K. Antosz, L. Knapčíková, and J. Husár, "Evaluation and Application of Machine Learning Techniques for Quality Improvement in Metal Product Manufacturing," *Applied Sciences (Switzerland)*, vol. 14, no. 10450, pp. 1–26, Nov. 2024, doi: 10.3390/app142210450.
15. F. H. Amrulloh, G. F. P. Aji, R. G. S, V. S. A. Anindyajati, R. N. Luthfi, and H. Tantyoko, "Klasifikasi Produktivitas Pekerja Garmen Menggunakan Algoritma Random Forest," *Jurnal Buffer Informatika*, vol. 10, no. 1, pp. 29–37, Apr. 2024, doi: 10.25134/buffer.v10i1.111.
16. N. Jourdan, S. Sen, E. J. Husom, E. Garcia-Ceja, T. Biegel, and J. Metternich, "On The Reliability Of Machine Learning Applications In Manufacturing Environments," *Workshop on Distribution Shifts, 35th Conference on Neural Information Processing Systems (NeurIPS 2021)*, Dec. 2021, [Online]. Available: http://arxiv.org/abs/2112.06986
17. I. D. Sulistyowati, S. Sunarno, and D. Djuniadi, "Penerapan Machine Learning Dengan Algoritma Support Vector Machine Untuk Prediksi Kelembapan Udara Rata-Rata," *Jurnal Sistem Informasi, Teknologi Informasi dan Komputer*, vol. 15, no. 1, pp. 234–324, Sep. 2024.
18. M. L. Pratama and H. Utama, "Pendekatan Deep Learning Menggunakan Metode LSTM Untuk Prediksi Harga Bitcoin," *IJCSR: The Indonesian Journal of Computer Science Research*, vol. 2, no. 2, 2023, doi: 10.37905.
19. F. Bayram and B. S. Ahmed, "A Domain-Region Based Evaluation of ML Performance Robustness to Covariate Shift," *Neural Comput. Appl.*, vol. 35, no. 24, pp. 17555–17577, May 2023, doi: 10.1007/s00521-023-08622-w.
20. M. R. Haas and L. Sibbald, "Measuring Data Drift with The Unstable Population Indicator," *Data Science*, vol. 7, pp. 1–12, Dec. 2024, doi: 10.3233/ds-240059.