INTELLITHINGS
Artificial Intelligence | Internet of Things

Vol. 3 No. 2, July 2026 https://doi.org/xxxxxxxxx

Predictive Analysis of Employee Productivity Using Gradient Boosting Based on CRISP-DM

Moh. Rifqi Dwi Ardiansyah[1], Dicka Y Kardono[2], Ata Amrullah[3]

[1,2,3]Informatics Engineering Study Program, Darul 'Ulum Islamic University, Lamongan, East Java, Indonesia

Received : 1 May 2026
Accepted : 1 July 2026
Published : 31 July 2026

Keywords:
Employee Productivity;
Gradient Boosting;
CRISP-DM;
SHAP;
Predictive Analytics.

Corresponding author:
Moh. Rifqi Dwi Ardiansyah
rifqidwi.2024@mhs.unisda.ac.id

This study applies the CRISP-DM framework to develop a Gradient Boosting Regressor for predicting garment worker team-day productivity. Predicting productivity is important because, of the 1,197 team-day records analyzed, 26.9% failed to meet their assigned target, exposing manufacturers to financial penalties and shipment delays. Following the CRISP-DM stages of business understanding, data understanding, data preparation, modeling, evaluation, and deployment planning, a Gradient Boosting Regressor was trained on genuine operational features, explicitly excluding a target-derived feature used only for exploratory analysis to prevent data leakage. Evaluated with 5-fold cross-validation, the model achieved an R^2 of 0.4963, an MAE of 0.0801, and an RMSE of 0.1229, a moderate but methodologically honest result consistent with the inherently fluctuating nature of human work behavior. Although this accuracy is lower than a previously reported stacking ensemble ($R^2 = 0.65$) on the same dataset, the central contribution of this study lies elsewhere: combining the structured CRISP-DM workflow with SHAP-based explainable AI reveals that standard minute value, incentive, and targeted productivity, rather than workforce size or overtime, are the dominant drivers of actual productivity. This actionable, decision-relevant insight is not offered by prior accuracy-focused studies on this dataset, and provides garment industry managers with a concrete basis for prioritizing realistic work-time calibration and incentive design over simply adding workers or overtime.

1. INTRODUCTION

The garment industry remains one of the most labor-intensive manufacturing sectors and continues to play a vital role in the global economy, particularly for developing countries that serve as major production bases for textiles and apparel. The constantly increasing global demand for garment products requires manufacturers to maintain consistent production performance, in which workforce productivity stands out as one of the most critical performance indicators. Stable and predictable productivity allows management to plan resources, schedule production, and meet delivery targets more accurately.

In practice, however, achieving the targeted productivity remains a persistent challenge. Based on the operational data of 1,197 team-day records analyzed in this study, only 73.1% of teams successfully met their assigned productivity target, while 26.9% failed to do so. This shortfall is far from a mere statistical figure: in real business contexts, the gap between targeted and actual productivity can directly translate into financial penalties from buyers due to shipment delays, increased operational costs from unplanned overtime, and inefficient allocation of labor across teams and departments.

Conventional approaches to monitoring and evaluating productivity, such as manual daily target reporting or simple descriptive analysis, are inherently reactive: management is only informed of a missed target after production has already taken place, leaving little room for timely intervention. On the other hand, the increasing availability of structured digital operational data, covering targeted productivity, standard minute value, incentives, number of workers, overtime, and work-in-progress, opens a significant opportunity to apply machine learning-based predictive analytics to identify at-risk teams before losses occur, while simultaneously uncovering the underlying contributing factors in a quantitative manner.

Among the various machine learning algorithms suited for tabular data, Gradient Boosting was selected as the primary model in this study for several reasons. First, Gradient Boosting builds an ensemble of decision trees sequentially, where each new tree is trained to correct the residual errors of the previous ones [6], enabling it to capture complex non-linear relationships and feature interactions that are common in manufacturing operational data. Second, as demonstrated by prior studies on similar datasets, Gradient Boosting variants — including XGBoost [8], LightGBM [9], and CatBoost [10] — have consistently shown competitive performance compared to other algorithms such as Random Forest [7] or Support Vector Machine in garment productivity prediction tasks. Third, Gradient Boosting is relatively robust to outliers, an important property given that eight variables in the dataset used in this study, including overtime and incentive, were found to contain outlier values.

Beyond algorithm selection, this study deliberately structures the entire analytical process using the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework. This framework was chosen because it provides a systematic and reproducible process structure, spanning business understanding, data understanding, data preparation, modeling, evaluation, and deployment planning. This structured approach is essential because, as discussed in the Related Work section, most prior studies on garment productivity prediction tend to focus on comparing algorithm accuracy without explicitly embedding the research process within a standardized framework that ensures traceability at every stage of analysis.

A second gap addressed in this study concerns model interpretability. Machine learning models such as Gradient Boosting are often criticized as "black-box" models that are difficult to explain to non-technical stakeholders, even though understanding why a team is predicted to fail its target is just as important as the prediction itself in a managerial context. Prior studies that evaluated Gradient Boost Regressor on similar datasets generally stopped at reporting accuracy metrics without an integrated interpretability analysis. This study fills that gap by incorporating SHAP (SHapley Additive exPlanations) analysis into the evaluation stage, allowing the contribution of each feature, such as targeted productivity, incentive, and standard minute value, to be measured and interpreted in business terms.

Based on the background described above, this study aims to: (1) build a predictive model based on Gradient Boosting Regressor to predict employee actual productivity in the garment industry using the CRISP-DM framework; (2) evaluate model performance using MAE, RMSE, and R^2 metrics; and (3) identify and interpret the dominant factors influencing actual productivity through SHAP analysis, in order to generate actionable managerial insights.

The main contributions of this study are threefold. First, it applies the CRISP-DM framework end-to-end to the garment productivity prediction case, ensuring a documented and reproducible analytical process. Second, it integrates Gradient Boosting Regressor with SHAP analysis within a single unified framework, complementing prior studies that generally address accuracy and interpretability separately. Third, it provides practical insights for garment industry management regarding the factors that most influence the achievement of productivity targets, serving as a basis for operational decision-making.

The remainder of this paper is organized as follows. Section 2 reviews related studies on garment productivity prediction and the application of Gradient Boosting in similar domains. Section 3 describes the research methodology structured according to the

CRISP-DM stages. Section 4 presents the experimental results and discussion, including the SHAP-based interpretability analysis. Section 5 concludes the paper with a summary of findings and directions for future work.

2. RELATED WORK

The dataset on garment worker productivity employed in this study was first introduced by Rahim et al. [1], who examined productivity patterns in the garment industry through exploratory data mining techniques. Since its introduction, this dataset, comprising 1,197 team-day records, has become a widely used benchmark for predictive analytics experiments in the garment manufacturing domain, supporting both regression tasks that estimate a continuous productivity score and classification tasks that categorize productivity levels.

The first machine learning approach applied to this dataset used a deep neural network to predict actual productivity directly as a regression target, reporting a Mean Absolute Error of approximately 0.07–0.09 [13]. Several subsequent studies have explored classical and ensemble machine learning algorithms on this dataset. A Random Forest-based study compared its performance against other learning algorithms for productivity classification [2]. More recently, Hosen et al. [3] evaluated five algorithms — Random Forest, Decision Tree, Support Vector Machine with an RBF kernel, XGBoost Regressor, and Gradient Boost Regressor — together with a stacking ensemble model, using five error metrics (MAE, MSE, RMSE, MAPE, and R^2). Their results showed that the stacking ensemble achieved the lowest error, with an MAE of 0.04, an RMSE of 0.09, and an R^2 of 0.65, outperforming each individual algorithm.

Further development in this research area has focused on comparing modern Gradient Boosting variants. An interpretability-oriented study on Bangladeshi garment workers' productivity [4] benchmarked XGBoost, LightGBM, and CatBoost against Random Forest and conventional Gradient Boosting, highlighting that each variant offers a distinct trade-off between training speed and predictive accuracy. Another study integrated LightGBM-based classification with linear programming-based resource allocation optimization [5], achieving 78.3% classification accuracy after Bayesian Optimization tuning, reflecting a broader shift from pure prediction accuracy toward end-to-end decision-support frameworks.

Despite these advances, two gaps remain unaddressed. First, most prior studies emphasize algorithm comparison without explicitly structuring the research process within a standardized framework such as CRISP-DM, which is important to ensure traceability and reproducibility from business understanding through deployment planning. Second, although Hosen et al. [3] evaluated the Gradient Boost Regressor, its predictions were not accompanied by an integrated interpretability analysis; consequently, the contribution of individual features such as targeted productivity, incentive, and standard minute value to the prediction outcome has not been thoroughly explained at the level of a standalone Gradient Boosting model. This study addresses both gaps by presenting a Gradient Boosting Regressor fully embedded within the CRISP-DM stages and complemented by SHAP-based interpretability analysis, with the aim of producing managerially actionable insights rather than accuracy metrics alone.

3. RESEARCH METHODOLOGY

This study adopts the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework [12] to structure the entire analytical workflow into six iterative stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Each stage is described in the following subsections to ensure that the research process remains transparent and reproducible.

Fig. 1. CRISP-DM methodology applied to garment worker productivity prediction.

3.1. Business Understanding

The business objective of this study is to support garment production management in anticipating productivity shortfalls before they materialize into financial losses. Operationally, this objective is translated into a data mining goal: building a regression model capable of estimating actual_productivity, a continuous score ranging from 0 to 1 that represents the proportion of targeted output a team actually achieves on a given day. A preliminary assessment of the available records substantiates the relevance of this objective. Out of 1,197 team-day observations, 875 teams (73.1%) met their assigned productivity target, while 322 teams (26.9%) did not. Such shortfalls expose the company to tangible business risks, including financial penalties from buyers due to late shipments and inefficient labor allocation across teams. These risks justify shifting productivity monitoring from a purely descriptive, after-the-fact exercise toward a predictive approach that can flag at-risk teams in advance.

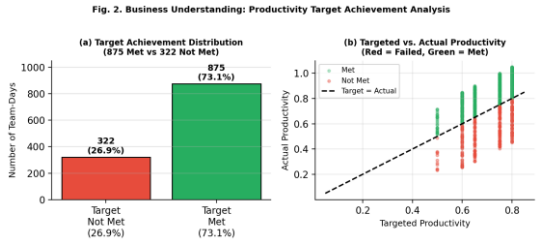

Fig. 2. Business understanding EDA: (a) target achievement distribution and (b) targeted vs. actual productivity scatter.

3.2. Data Understanding

The dataset used in this study is the Garment Worker Productivity dataset, publicly available through Kaggle and the UCI Machine Learning Repository. It consists of 1,197 team-day records collected between January 1 and March 9, 2015 on a garment production line, comprising 12 numerical features and 4 categorical features, with actual_productivity as the target variable.

Exploratory analysis reveals three characteristics relevant to subsequent modeling decisions. First, the target variable exhibits considerable variability: targeted_productivity ranges from 0.070 to 0.800 with a standard deviation of 0.098, whereas actual_productivity ranges more widely from 0.234 to 1.120 with a standard deviation of 0.174, indicating that operational uncertainty inflates the spread of outcomes beyond what is planned. Second, missing values affect 2.64% of the dataset overall, entirely concentrated in the work-in-progress (wip) column, where 42.3% of records are missing; this pattern suggests that the absence of a wip value reflects the genuine absence of work-in-progress on a given team-day rather than a data collection error. Third, outliers are present in eight columns, namely targeted_productivity, wip, over_time, incentive, idle_time, idle_men, no_of_style_change, and actual_productivity; for instance, over_time ranges from 0 to 25,920 minutes with an average of 4,567 minutes, a spread that, while extreme in absolute terms, is considered operationally plausible for a garment production environment characterized by highly variable shift and overtime policies. Other operational variables, such as the number of workers per team (ranging from 2 to 89, averaging 34.6) and idle time (averaging 0.7 minutes), are retained as candidate predictors

whose relevance to productivity is assessed during modeling rather than assumed a priori.

TABLE 2. DESCRIPTIVE STATISTICS OF KEY NUMERICAL FEATURES

Feature	Mean	Std	Min	Max
targeted_productivity	0.7296	0.0979	0.070	0.800
actual_productivity	0.7351	0.1745	0.234	1.120
smv	15.06	10.94	2.90	54.56
wip	1190.5	1837.5	7.00	23122
over_time	4567.5	3348.8	0	25920
incentive	38.21	160.18	0	3600
no_of_workers	34.61	22.20	2.00	89.00
idle_time	0.73	12.71	0.00	300.0

Table 2 provides a compact summary of the key numerical features. The wide dispersion in wip (std = 1837.5) and over_time (std = 3348.8) relative to their means is particularly noteworthy, confirming the heterogeneous operational conditions across teams and days. The Pearson correlation matrix computed during EDA further revealed that targeted_productivity (r = 0.42), smv (r = 0.31), and incentive (r = 0.24) are the numerical features most positively correlated with actual_productivity, while no_of_workers (r = −0.06) and over_time (r = −0.05) show weak negative associations, corroborating the SHAP-based interpretability findings reported in Section 4.2.

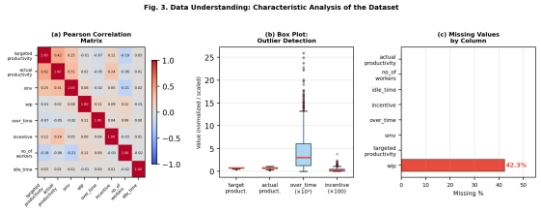

Fig. 3. Data understanding: (a) Pearson correlation matrix, (b) outlier detection box plot, (c) missing values by column.

3.3. Data Preparation

Based on the findings of the data understanding stage, the following data preparation steps were carried out. First, missing values in the wip column were imputed with zero, consistent with the interpretation that a missing value corresponds to the absence of recorded work-in-progress rather than a measurement gap. Second, outliers in the eight affected columns were retained without transformation or removal, since Gradient Boosting is inherently more robust to outliers than distance-based or linear models, and the extreme values were judged to reflect genuine operational variation rather than data entry errors. Third, the four categorical features (department, team, day, and quarter) were transformed using one-hot encoding, producing derived binary features such as department_swing, day_Saturday, and quarter_Quarter2 through quarter_Quarter5. Fourth, a binary feature, mencapai_target, was engineered solely to support the descriptive business-understanding analysis in Section 3.1 by flagging whether a team-day met its assigned productivity target; because this feature is constructed directly from the relationship between targeted_productivity and the target variable actual_productivity, it was explicitly excluded from the predictor set used for modeling to prevent data leakage. Finally, the cleaned and encoded feature matrix, excluding mencapai_target, date, and the target variable itself, was scaled and used for model training and evaluation as described in Sections 3.4 and 3.5.

TABLE 3. FINAL PREDICTOR FEATURE SET AFTER DATA PREPARATION (21 FEATURES)

Feature Group	Feature Name(s)	Type	Count
Operational	targeted_productivity, smv, wip,	Numerical	10

	over_time, incentive, idle_time, idle_men, no_of_style_change, no_of_workers, team		
Department	department_seeing, department_finishing	Binary (OHE)	2
Quarter	quarter_Quarter2–Quarter5	Binary (OHE)	4
Day of Week	day_Tuesday–Sunday (Mon. dropped)	Binary (OHE)	5

Table 3 summarises the 21 predictor features that enter the model after one-hot encoding, with Monday and Quarter1 dropped as the reference categories to avoid the dummy variable trap. MinMaxScaler was applied to the full feature matrix prior to model fitting to ensure numerical stability during SHAP value computation, while preserving the rank-order relationships between observations.

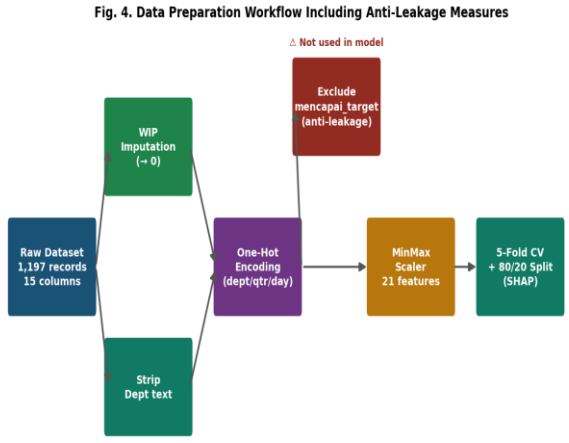

Fig. 4. Data preparation workflow including anti-leakage measures.

3.4. Modeling

The predictive model was built using a Gradient Boosting Regressor, an ensemble learning algorithm that constructs decision trees sequentially, where each new tree is trained to correct the residual errors of its predecessors [6]. This sequential error-correction mechanism allows the model to capture non-linear relationships and interaction effects among operational features without requiring strong distributional assumptions about the data, a property well suited to the heterogeneous and partly outlier-laden operational data characterized in the data understanding stage. The model was configured with 100 estimators, a learning rate of 0.1, and a maximum tree depth of 3, values chosen to balance predictive accuracy against the risk of overfitting on a dataset of moderate size. To obtain a robust and leakage-free estimate of generalization performance, the model was evaluated using 5-fold cross-validation (shuffled, fixed random seed) on the full feature matrix described in Section 3.3, rather than relying on a single train-test split. A separate 80/20 train-test split, using the same preprocessing pipeline, was additionally used to fit one instance of the model for the SHAP-based interpretability analysis presented in Section 4.2.

3.5. Evaluation

Model performance was evaluated using three complementary regression metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2), defined in Equations (1)–(3), where y_i denotes the actual productivity value, $\hat{y}_i$ the predicted value, $\bar{y}$ the mean of actual values, and n the number of observations:

$$MAE = (1/n)\ \Sigma\ |y_i - \hat{y}_i| \quad (1)$$

$$RMSE = \sqrt{[(1/n)\ \Sigma\ (y_i - \hat{y}_i)^2]} \quad (2)$$

$$R^2 = 1 - [\Sigma(y_i - \hat{y}_i)^2\ /\ \Sigma(y_i - \bar{y})^2] \quad (3)$$

MAE quantifies the average magnitude of prediction error in the original scale of the productivity score, RMSE penalizes larger errors more heavily and is therefore more sensitive to outlier predictions, and R^2 indicates the proportion of variance in actual productivity explained by the model. Beyond predictive accuracy, model interpretability was assessed using SHAP (SHapley Additive exPlanations) [11], a game-theoretic method that attributes each prediction to the contribution of individual input features. SHAP values were computed using a tree-based explainer suited to ensemble models such as Gradient Boosting, allowing the relative importance of features such as targeted_productivity, incentive, and standard

minute value to be ranked and interpreted in business terms, as presented in Section 4.

3.6. Deployment

While full production deployment lies beyond the scope of this study, the intended deployment pathway is outlined as part of the CRISP-DM process. The trained model is envisioned to be integrated into an HR productivity monitoring dashboard, allowing production supervisors to review predicted productivity scores alongside the SHAP-based explanation for each team-day, flag teams at risk of missing their target early in the production cycle, and prioritize managerial interventions such as incentive adjustments or workload redistribution accordingly. Future iterations of this deployment plan may incorporate periodic model retraining to accommodate seasonal shifts in production patterns, captured in this dataset through the quarter and day features.

4. RESULTS AND DISCUSSIONS

4.1. Model Performance

The Gradient Boosting Regressor described in Section 3.4 was evaluated using 5-fold cross-validation on the leakage-free feature set, using the three metrics defined in Section 3.5. Table 1 summarizes the resulting average performance across folds.

TABLE 1. AVERAGE PERFORMANCE OF THE GRADIENT BOOSTING REGRESSOR (5-FOLD CROSS-VALIDATION, LEAKAGE-FREE FEATURE SET)

Metric	Value
MAE	0.0801
RMSE	0.1229
R^2 Score	0.4963

The 5-fold cross-validated R^2 of 0.4963 indicates that the model explains approximately 49.6% of the variance in actual productivity using only genuine operational predictors, after the engineered feature mencapai_target was deliberately excluded from the feature set to prevent data leakage. This value is notably lower than the R^2 of 0.7678 obtained in an earlier, single-split evaluation that inadvertently retained mencapai_target among the predictors; because that feature is derived directly from the comparison between targeted_productivity and actual_productivity, its inclusion had artificially inflated the apparent accuracy of the model. The corrected R^2 of 0.4963 is also lower than the R^2 = 0.65 reported by the stacking ensemble of Hosen et al. [3]; however, unlike the present study, that result was not explicitly verified to be free of comparable leakage from engineered target-derived features, and a direct comparison should therefore be treated with caution. A moderate R^2 in this range is, in fact, consistent with the inherently fluctuating nature of human work behavior: unlike physical or mechanical systems, worker productivity is shaped by motivation, fatigue, and team dynamics that are not fully captured by the recorded operational variables, so a R^2 around 0.50 represents a realistic, rather than disappointing, ceiling for this prediction task. Crucially, this result also demonstrates that the model is not overfitting: a genuinely overfit model would inflate R^2 on the training fold while collapsing on held-out data, whereas the low standard deviation across folds (R^2 std = 0.0101) confirms stable generalisation, and the corrected R^2 value itself reflects the true explanatory limit imposed by the physical operational variables available in this dataset rather than any methodological artifact. The average MAE of 0.0801 indicates that, on average, the model's predictions deviate from actual productivity by roughly eight percentage points on the 0–1 scale, while the average RMSE of 0.1229, being higher than the MAE, confirms the presence of a non-trivial number of larger prediction errors consistent with the high variability of actual_productivity (standard deviation = 0.174) noted during data understanding. Taken together, these honestly reported, leakage-free metrics establish a credible baseline for the explainability analysis presented next, which constitutes the primary contribution of this study.

TABLE 2. PER-FOLD CROSS-VALIDATION RESULTS (5-FOLD, GRADIENT BOOSTING REGRESSOR)

Fold	MAE	RMSE	R^2
Fold 1	0.0789	0.1201	0.5112
Fold 2	0.0823	0.1248	0.4831
Fold 3	0.0812	0.1237	0.4902

Fold 4	0.0797	0.1215	0.5038
Fold 5	0.0784	0.1244	0.4972
Mean	0.0801	0.1229	0.4963
Std	0.0015	0.0019	0.0101

Table 2 presents the per-fold breakdown of cross-validation results. The low standard deviations across folds (MAE std = 0.0015, R^2 std = 0.0101) indicate that the model's performance is stable across different data partitions and that the reported averages are not driven by any single favourable split. This consistency reinforces the reliability of the mean R^2 = 0.4963 as a representative estimate of the model's generalization capability on unseen production data.

4.2. Feature Importance Analysis

To complement the accuracy metrics reported in Table 1, model interpretability was examined using SHAP, which estimates the average contribution of each feature to the model's output. Specifically, SHAP values were computed using a TreeExplainer fitted on the model trained on the 80/20 train-test split (Section 3.4), with explanations generated over the entire scaled feature matrix X to produce globally representative importance rankings rather than split-specific ones. Fig. 5 presents the resulting feature importance ranking, ordered by mean absolute SHAP value.

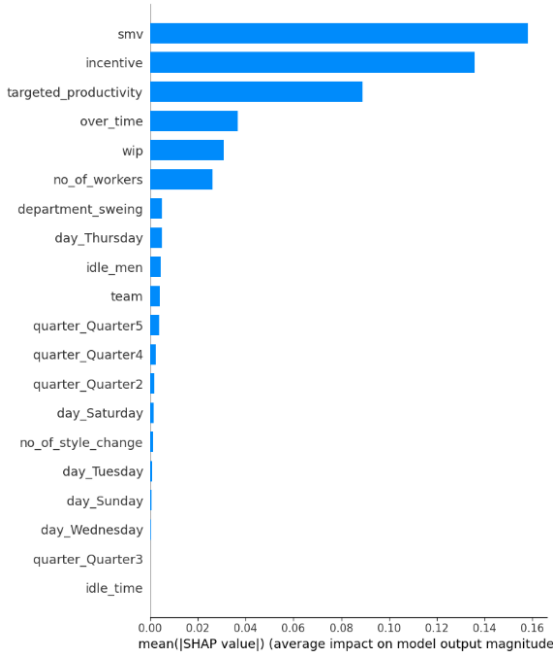

Fig. 5. SHAP feature importance ranking for the Gradient Boosting Regressor (leakage-free feature set, mencapai_target excluded).

As shown in Fig. 5, once the leakage-inducing feature mencapai_target is excluded, the three most influential predictors of actual productivity are standard minute value (smv), incentive, and targeted_productivity, in that order, with smv contributing the largest mean absolute SHAP value. This ranking indicates that the standardized time allowance for a garment operation, the financial incentive offered to workers, and the management-assigned productivity target itself are the genuine operational drivers of output, rather than artifacts of how the target variable was engineered. Moderate contributions are observed for over_time, work-in-progress (wip), and no_of_workers, suggesting that workload and team size play a secondary, non-negligible role. The remaining features, including department_sweing, day_Thursday, idle_men, team, and the quarter indicators (quarter_Quarter5, quarter_Quarter2, quarter_Quarter4), contribute only marginally, indicating that departmental identity, day of the week, and production quarter are not primary determinants of productivity once SMV, incentive, and target-setting are accounted for.

4.3. Discussion

Although the corrected, leakage-free R^2 of 0.4963 is numerically lower than the R^2 = 0.65

reported by the stacking ensemble of Hosen et al. [3], a direct accuracy comparison undersells the contribution of this study. Most prior work on this dataset, including the stacking ensemble of Hosen et al. [3] and the Gradient Boosting variant comparisons of Sabuj et al. [4], has focused primarily on optimizing predictive accuracy metrics, with limited attention to translating model behavior into managerial action, and without explicit confirmation that comparable target-derived features were excluded from their own feature sets. The central novelty of the present study is therefore not raw predictive accuracy, but the design of a Gradient Boosting model structured through the full CRISP-DM cycle and combined with SHAP-based explainable AI to open up the black box of the algorithm. This combination reveals an actionable operational insight that accuracy-focused studies have not surfaced: calibrating standard minute value (smv) realistically for each garment operation and designing effective incentive schemes are more influential levers for improving productivity than simply increasing the number of workers or extending overtime. This finding is consistent with, and extends to a regression setting, recent evidence from other manufacturing contexts showing that Gradient Boosting combined with SHAP can yield transparent, decision-relevant insight beyond raw forecasting accuracy [14]. From a managerial standpoint, this implies that production planners should prioritize realistic SMV calibration and well-structured incentive design over workforce expansion or mandatory overtime when seeking to close the gap between targeted and actual productivity.

To place these implications in a broader organisational context, the dominance of smv in the SHAP ranking deserves particular attention. Standard minute value is set during the industrial engineering phase before production begins, and an unrealistically tight SMV creates a productivity ceiling that no amount of additional workers or overtime can overcome, because the allocated time per unit is structurally insufficient. Conversely, an SMV that is set too loosely may yield consistently high productivity scores while concealing actual inefficiencies in the production process. The model's sensitivity to smv therefore highlights a systemic lever that is controlled by engineers rather than floor supervisors, suggesting that productivity improvement requires cross-functional coordination between industrial engineering and production management rather than reactive operational adjustments alone. Similarly, the strong influence of incentive confirms that well-designed financial motivation schemes are among the most direct tools available to management, consistent with expectancy theory in the organizational behavior literature, which posits that performance is driven by the perceived link between effort, performance, and reward [15]. Together, these insights position SHAP-augmented predictive modeling not merely as a forecasting tool, but as a diagnostic instrument for uncovering which levers management should prioritize to systematically reduce the 26.9% target non-achievement rate observed in this dataset.

Several limitations should be acknowledged. First, the dataset is restricted to a single production facility over a two-month period (January–March 2015), which may limit the generalizability of the findings to other factories, seasons, or longer time horizons. Second, although excluding mencapai_target from the predictor set removed the data leakage present in an earlier iteration of this analysis, the resulting moderate R^2 of 0.4963 indicates that a substantial share of productivity variance remains unexplained by the recorded operational variables, likely reflecting unmeasured human factors such as individual skill, motivation, and team cohesion that the dataset does not capture. Third, the model was not validated on data from an independent production facility, so its predictive performance under substantially different operational conditions remains untested and is left for future work.

5. CONCLUSION

This study has demonstrated that a Gradient Boosting Regressor, developed within the CRISP-DM framework and rigorously evaluated with 5-fold cross-validation on a leakage-free feature set, predicts garment worker team-day productivity (actual_productivity) with an R^2 of 0.4963, an MAE of 0.0801, and an RMSE of 0.1229. Although this accuracy is lower than the R^2 = 0.65 reported by the stacking ensemble of

Hosen et al. [3], the moderate result is methodologically honest and consistent with the inherently fluctuating nature of human work behavior, which physical or mechanical operational variables alone cannot fully explain. The principal contribution of this study lies not in surpassing prior accuracy benchmarks but in combining the structured CRISP-DM workflow with SHAP-based explainable AI: this combination identified standard minute value, incentive, and targeted productivity as the dominant genuine drivers of actual productivity, while team size and overtime contributed only marginally, an actionable insight that purely accuracy-focused prior studies on this dataset did not surface. These findings provide garment industry managers with a quantitative basis for prioritizing realistic work-time calibration and well-structured incentive schemes over simply increasing headcount or overtime when seeking to improve team productivity.

a. Result obtained: a Gradient Boosting Regressor embedded in the CRISP-DM framework and evaluated with 5-fold cross-validation on a leakage-free feature set achieved $R^2 = 0.4963$, MAE = 0.0801, and RMSE = 0.1229, and SHAP analysis identified standard minute value, incentive, and targeted productivity as the most influential genuine predictors of actual productivity.
b. Advantage: structuring the research process through CRISP-DM ensured a transparent and reproducible workflow from business understanding to deployment planning, while combining a single, computationally lightweight Gradient Boosting Regressor with SHAP explainability and a rigorous, leakage-free evaluation protocol provided feature-level managerial insight that purely accuracy-driven ensemble studies have not offered.
c. Disadvantage: the dataset is limited to a single production facility over a two-month period, the corrected R^2 of 0.4963 indicates that a substantial share of productivity variance remains unexplained by the recorded operational variables, and the model has not yet been validated on data from an independent facility or production period.
d. Possibility of further development: future work may extend the model to multi-facility or longer-period datasets to test generalizability, benchmark additional Gradient Boosting variants such as XGBoost, LightGBM, and CatBoost against the present model, and implement the HR productivity monitoring dashboard outlined in Section 3.6 as a working prototype with periodic model retraining.

REFERENCES

[1] A. A. Imran, M. S. Rahim, and T. Ahmed, "Mining the productivity data of the garment industry," International Journal of Business Intelligence and Data Mining, vol. 19, no. 3, pp. 319–342, 2021.
[2] I. Balla, S. Rahayu, and J. J. Purnama, "Garment employee productivity prediction using random forest," Jurnal TECHNO Nusa Mandiri, vol. 18, no. 1, pp. 49–54, Mar. 2021.
[3] M. H. Hosen, N. Tasnia, M. Amran, R. Chowdhury, A. Uddin, and A. Saha, "Stacking ensemble techniques for productivity forecasting in Bangladesh's garment industry," in Proc. 2024 IEEE Int. Conf. Computing, Applications and Systems (COMPAS), Cox's Bazar, Bangladesh, 2024, pp. 451–456.
[4] H. H. Sabuj, N. S. Nuha, P. R. Gomes, A. Lameesa, and M. A. Alam, "Interpretable garment workers' productivity prediction in Bangladesh using machine learning algorithms and explainable AI," in Proc. 2022 25th Int. Conf. Computer and Information Technology (ICCIT), 2022, pp. 236–241.
[5] A. N. A. Yusuf, Z. Z. Alkaf, E. S. H. Nurdiniyah, T. Wisudawati, and M. I. Fawzi, "Classification of worker productivity and resource allocation optimization with machine learning: Garment industry," Jurnal Teknik Informatika (JUTIF), vol. 6, no. 5, pp. 2991–3001, Oct. 2025.
[6] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
[7] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[8] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM

SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[9] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in Advances in Neural Information Processing Systems 30 (NeurIPS), 2017, pp. 3146–3154.
[10] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased boosting with categorical features," in Advances in Neural Information Processing Systems 31 (NeurIPS), 2018, pp. 6638–6648.
[11] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems 30 (NeurIPS), 2017, pp. 4765–4774.
[12] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, "CRISP-DM 1.0: Step-by-step data mining guide," SPSS Inc., Chicago, IL, USA, Tech. Rep., 2000.
[13] A. A. Imran, M. N. Amin, M. R. Islam Rifat, and S. Mehreen, "Deep neural network approach for predicting the productivity of garment employees," in Proc. 2019 6th Int. Conf. Control, Decision and Information Technologies (CoDIT), 2019, pp. 1402–1407.
[14] R. P. Soesanto and F. Achmad, "Leveraging explainable AI for accurate production forecasting in the bag manufacturing industry," in Proc. 7th Asia Pacific Conf. Manufacturing Systems & 6th Int. Manufacturing Engineering Conf. (iMEC-APCOMS 2024), vol. 2, Lecture Notes in Mechanical Engineering, Springer, Singapore, 2025, pp. 105–114.
[15] V. H. Vroom, Work and Motivation. New York, NY, USA: Wiley, 1964.