DIJEMSS: Dinasti International Journal of Education Management and Social Science
E-ISSN: 2686-6331
P-ISSN: 2686-6358
https://dinastipub.org/DIJEMSS dinasti.info@gmail.com +62 811 7404 455
DINASTI PUBLISHER
DOI: https://doi.org/10.38035/dijemss.v6i5
https://creativecommons.org/licenses/by/4.0/

Logistics Management Optimization through Machine Learning: A Predictive Model for Item Transfer Time in Warehouse Activity-Space

Hendri Lasmana[1], Agus Purnomo[2], Erna Mulyati[3]
[1]Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia, hendriboilasmana@gmail.com
[2]Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia, aguspurnomo@ulbi.ac.id
[3]Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia, ernamulyati@ulbi.ac.id

Corresponding Author: hendriboilasmana@gmail.com[1]

Abstract: Operational efficiency in warehouse logistics relies heavily on accurately predicting item transfer time. This study presents a machine learning-based framework using Gradient Boosting Classifier to classify transfer durations in the dynamic Jakarta Centrum warehouse, managed by the Corruption Eradication Commission (KPK) and PosIND. Field observations revealed inefficiencies due to unstructured layouts and fluctuating volumes. To improve prediction accuracy, the model incorporates Z-score normalization, SMOTE for class balancing, and hyperparameter tuning using GridSearchCV and PSO. The optimized model successfully classified 258 High, 285 Low, and 277 Medium transfer-time instances. SHAP analysis identified distance, distribution volume, and throughput as key influencing factors. Results demonstrate the potential of predictive modeling to enhance warehouse operations through better space usage, workforce planning, and SLA compliance. This study supports machine learning as a strategic tool for data-driven logistics optimization, with future work recommended to include contextual variables like workforce capacity and shift schedules for improved precision and real-world applicability.

Keyword: Logistics management, item transfer time, warehouse optimization, machine learning, gradient boost, hyperparameter tuning, SHAP, SMOTE.

INTRODUCTION

The optimization of hyperparameter tuning holds substantial promise in enhancing the accuracy of item transfer time predictions within activity-space-based logistics scenarios. This research addresses a key challenge in logistics management, namely improving warehouse operational efficiency by accurately predicting the duration of item movement. The study integrates machine learning as a decision support mechanism to complement logistic strategies related to space utilization, picking operations, and throughput balance. The research is situated in a dynamic, high-complexity environment: specifically, the warehouse operations of the Corruption Eradication Commission (KPK) in collaboration with PosIND at the Jakarta Centrum facility. Field observations at this site revealed significant operational inefficiencies, including unstructured item categorization, disorganized layouts, and suboptimal workflow processes—factors that directly contribute to increased transfer times and degraded service levels.

Instead of positioning machine learning as the primary novelty, this study employs it as a supportive methodology within a logistics-centric framework. The predictive model aims to facilitate real-time estimations that allow warehouse managers to proactively reorganize layouts, reallocate resources, and adjust scheduling based on forecasted transfer time categories. Through this, the research seeks to improve key logistics performance indicators such as service level agreement (SLA) compliance, order cycle time, and labor productivity.

Traditional warehouse layout strategies such as Dedicated Storage and Class-Based Storage have long been utilized in inventory management. However, in dynamic warehouse environments marked by fluctuating inventory volumes and unpredictable item demands, these methods are becoming increasingly insufficient. Dedicated storage assigns fixed locations to items, which leads to poor space utilization when inventory fluctuates. Meanwhile, class-based storage—though slightly more flexible—relies on static assumptions about demand patterns, which often do not align with real-world dynamics. As a result, these traditional strategies fail to accommodate the evolving complexities of modern activity-space environments and limit predictive real-time decision-making. The machine learning-based predictive modeling offers a more adaptive and data-driven solution for estimating item transfer times in real time. Models optimized through hyperparameter tuning can dynamically adjust to changing distribution patterns, item volumes, and levels of operational urgency. The importance of hyperparameter optimization in machine learning has been widely recognized, where appropriate tuning significantly impacts model performance (Ilemobayo et al., 2024; Ramadhan & Pane, 2024; Xiong et al., 2020).

Recent studies have emphasized the rising role of machine learning in logistics optimization, particularly in forecasting operational metrics like item transfer time within warehouse activity space. In this study, Particle Swarm Optimization (PSO) was utilized to fine-tune the hyperparameters of a gradient-boosting classifier, leveraging swarm intelligence for efficient exploration and delivering superior convergence and predictive outcomes (W. Zhang & Li, 2023). To handle class imbalance within operational data, the SMOTE-Tomek links method was applied—an effective approach to class balancing that also reduces boundary overlap and enhances generalization (Liu & Chen, 2022). Furthermore, model explainability was achieved using SHapley Additive exPlanations (SHAP), which helped identify key features influencing transfer time and offered reliable information for decision making (Kumar & Singh, 2023).

Model complexity and data quality, when combined with effective hyperparameter tuning, can significantly improve the performance of predictive algorithms (Ilemobayo et al., 2024). A comprehensive model development process must systematically explore diverse tuning strategies, including both simple methods like Grid Search and more advanced techniques as outlined by Bischl et al. (Bischl et al., 2023). These strategies represent evolving practices for efficient model optimization.

The critical importance of hyperparameter optimization extends beyond algorithm-level enhancements and directly impacts the robustness of predictive models as a whole. Studies have shown that automated and adaptive tuning techniques such as those developed by Iqbal (Iqbal et al., 2022) and Bahmani (Bahmani et al., 2021) simplify the tuning process while prioritizing influential configurations within large parameter spaces. Additionally, adaptive optimizers as described by Mohapatra (Mohapatra et al., 2021) demonstrate promising results in tailoring hyperparameter settings to the specific requirements of models in various domains.

The remainder of this paper is organized as follows. Section Research Design describes the research design, including data collection, pre-processing, modeling approach, and hyperparameter tuning strategy. Section Results And Discussion presents the model evaluation results, discusses key findings and limitations, and compares them with related works to position the study within existing research. Finally, Section Conclusion concludes the study and outlines potential directions for future research.

METHOD

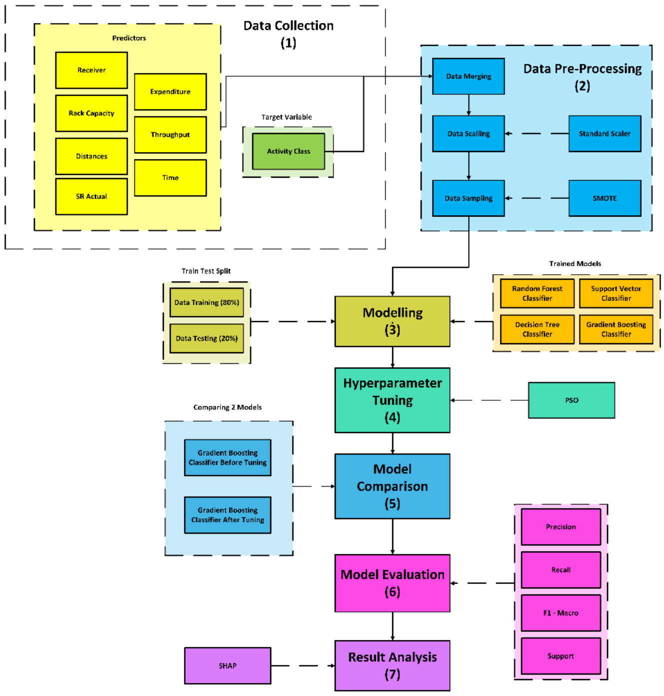

Figure 2. 1 The proposed research methodology.

A. Data Collection

The research began by collecting 1,000 daily operational records from an activity-based logistics system, which were stored in two separate Excel files. These files were merged using a common identifier to create a unified data set suitable for analysis. The data set included seven independent variables: Receiving, Distribution, Rack Capacity, Throughput, Distance, Time, and Actual Service Level, which are categorized into independent variables (X). The dependent variable (Y) is the Activity Class, which categorizes each record based on the time it takes to move the goods.

These variables were chosen because structural and operational factors, such as distance and rack capacity, are known to significantly affect the estimated transfer time (Khiari & Olaverri-Monreal, 2020). The variables were systematically collected to reflect the complexity of real-world activity-space scenarios. Before proceeding to the modeling and hyperparameter tuning stages, the dataset underwent important pre-processing steps, including merging, structuring, and target class segmentation, to ensure consistency, completeness, and readiness for predictive analysis.

B. Pre-Processing

After categorizing data into independent and dependent variables, a thorough pre-processing phase was performed to transform raw inputs from IoT sensors and warehouse systems into machine learning-ready formats. This step was crucial for revealing patterns in item transfers and ensuring model reliability (Brijith, 2023). Data from multiple sources were merged using activity IDs and synchronized timestamps to maintain consistency and contextual alignment, which is essential for effective learning (Su et al., 2024).

Next, feature scaling was performed to standardize the different measurement units in features such as distance (in meters), time (in seconds), throughput (units / hour) and rack capacity. Z-score normalization was applied:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

Formula 1 standardizes features using the mean (μ) and standard deviation (σ), setting each feature to zero mean and unit variance to enhance model stability, especially in gradient-based classifiers like Gradient Boosting (Fan et al., 2022). The dataset, categorized into "high," "medium," and "low" transfer times, was imbalanced with "medium" as the dominant class. To address this, SMOTE was used to generate synthetic samples for the minority classes, improving recall and reducing bias—an approach shown to be effective in logistics classification tasks (Y. Zhang et al., 2024).

C. Modeling

After pre-processing steps such as merging, normalization, and balancing, the Gradient Boosting Classifier was used to predict item transfer time based on activity space characteristics. As a sequential ensemble method, it effectively captures complex non-linear relationships in structured, imbalanced datasets common in warehouse environments (Derby et al., 2023). Compared to traditional models like logistic regression and decision trees, Gradient Boosting shows superior performance on metrics such as F1 score and AUC, especially with structured data (Song et al., 2023). The model was implemented using scikit-learn's Gradient Boosting Classifier, initially with default hyperparameters for baseline evaluation. To ensure robust performance evaluation, the dataset was split into training and testing sets, and further validated using k-fold cross-validation. This approach reduces variance in performance estimates and prevents overfitting by assessing the model across multiple data folds. Although often considered a "black box," the interpretability of Gradient Boosting was enhanced using SHAP, enabling feature-level insights that support practical decision-making in logistics contexts (Salih et al., 2024).

D. Hyperparameter Tuning

Following the initial modeling using the Gradient Boosting Classifier, the next essential phase involves hyperparameter tuning. This process is carried out to determine the most suitable combination of model parameters that maximizes classification performance to predict item transfer time in activity-based logistic scenarios.

In this study, hyperparameter tuning is applied specifically to improve the precision of the Gradient Boosting Classifier model in classifying transfer-time categories such as 'high','medium', and 'low.' Default hyperparameter settings are often inadequate for handling the complexity and unbalanced nature of logistic data, particularly when rare classes are involved (Boldini et al., 2022). PSO is applied in the context of hyperparameter tuning, where it explores combinations of parameters through cross-validation(Pane et al., 2022). This approach has proven highly effective in improving model performance for logistic prediction problems, particularly those affected by class imbalance (Kocsis Szürke et al., 2022). Following the tuning process, the optimized model is evaluated against a baseline to determine improvements in metrics such as F1 macro and recall. This method contributes to better generalization and more reliable predictions, which are crucial to maintain accuracy and operational reliability in real-world logistics scenarios (Tang et al., 2023).

E. Evaluation Metrics

To evaluate the multiclass model that predicts item transfer time, we used a suite of standard metrics tailored for unbalanced datasets and multiclass classification (Farhadpour et al., 2024):

- **Precision -** the proportion of correctly predicted positive instances out of all predicted positives:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Formula 2 defines precision as the ratio of true positive predictions (**TP**) to the total predicted positive instances, which includes both true positives and false positives (**TP + FP**). Here, **TP** denotes the number of correctly identified positive cases, while **FP** refers to instances incorrectly classified as positive. This metric reflects the model's ability to avoid false alarms and measures the accuracy of positive predictions (Saito & Rehmsmeier, 2020).

- **Recall -** the proportion of true positives captured out of all actual positives:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

Formula 3 measures the ratio of true positives (**TP**) to all actual positives (**TP + FN**), where **FN** represents false negatives. It indicates the model's ability to correctly identify all relevant instances of a class (Farhadpour et al., 2024).

- **F1-Score -** the harmonic mean of precision and recall:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Formula 4 calculates the harmonic mean between precision and recall, where **precision** is the ratio of true positives to all predicted positives, and **recall** is the ratio of true positives to all actual positives. This metric provides a balanced evaluation considering both false positives and false negatives, making it particularly effective in handling class imbalance (Opitz, 2024; Saito & Rehmsmeier, 2020).

- **Support -** the number of actual instances per class; it contextualizes other metrics and influences weighted averaging (Farhadpour et al., 2024).

Together, these metrics ensure a robust, fair and detailed evaluation of model performance, particularly critical in our imbalanced, multiclass scenario.

F. EXPLANABILITY WITH SHAP

This study utilized SHAP to interpret the contributions of the characteristics in the Gradient Boosting Classifier. As a game-theoretic model-agnostic approach, SHAP assigns an importance value to each feature by quantifying its marginal contribution to the model output (Lundberg & Lee, 2020). By aggregating SHAP values, the analysis revealed which features most influenced classification decisions across transfer-time categories. This approach not only improves global interpretability, but also supports consistent class-level explanations of characteristic behavior (Tang et al., 2023).The integration of SHAP in this study ensured transparency in how the model arrived at its predictions, addressing the common black-box nature of ensemble methods while allowing for more informed operational insights.

RESULTS AND DISCUSSION

A. Result

The results of this study are presented in sequential order, from data preparation to interpretation of the model output. Each stage plays a critical role in ensuring the accuracy and reliability of the final predictive model.

During preprocessing, normalization and SMOTE were applied to improve data consistency and address class imbalance - factors essential for robust classification performance (de Amorim et al., 2022). Feature normalization was applied to bring all variables to a comparable scale, while SMOTE balanced the class distribution, resulting in improved recognition of minority categories.

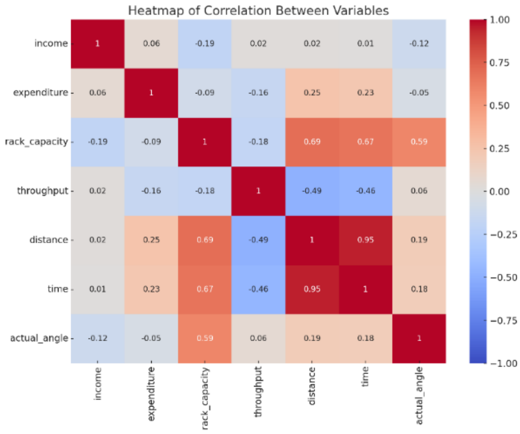

Figure 2. 2 Heatmap showing correlation between operational and

Pearson's correlation analysis was used at an early stage to assess the relationship between features in the data set. As shown in Figure 2.2 the analysis reveals a very high correlation between distance and time (**r = 0.95**), as well as a moderate correlation with rack capacity (distance: **r = 0.69**, time: **r = 0.67**). The capacity of the rack is also moderately correlated with actual service levels (**r = 0.59**). Although distance and time exhibit linear redundancy, they are still included in the model, as they represent different operational dimensions and have been shown to strengthen predictive performance. Therefore, the final features used include distance, time, throughput, rack capacity, and actual service level.

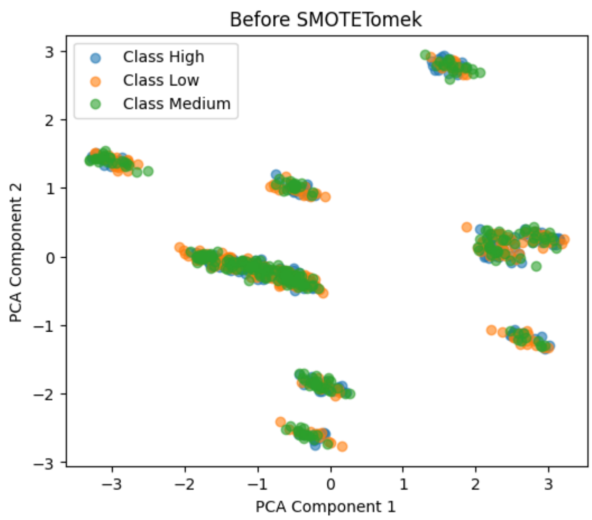

Figure 2. 3 PCA visualization before applying SMOTE

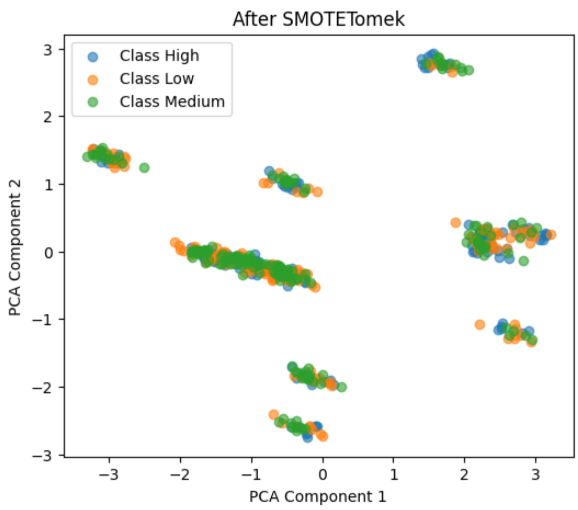

Figure 2. 4 PCA visualization after applying SMOTE.

Furthermore, to handle the identified class imbalance, the SMOTE technique was applied both before and after data transformation using principal component analysis (PCA). An initial visualization of the PCA results Figure 2.3 shows that the class distribution is unbalanced, with the 'Medium' class dominating 435 samples, while the ‘High’ and ‘Low’ classes are much lower, only 67 and 78 samples, respectively. This imbalance leads to overlapping groups and unclear boundaries between classes. Following the application of the SMOTE Figure 2.4, the data set achieves a more uniform class distribution, with 285 samples assigned to each class. The updated PCA projection exhibits enhanced class separation and a more balanced dispersion throughout the feature space. These results demonstrate the efficacy of the balancing strategy in improving data representation, thus facilitating more effective and generalizable model training.

Figure 2. 5 Scatter plot of actual item transfer time vs. distance based on activity class

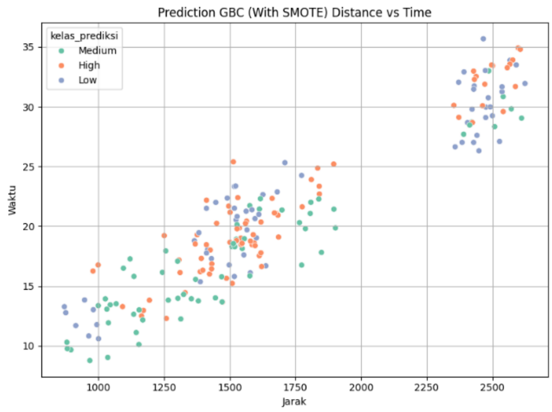

Figure 2. 6 Scatter plot of model-predicted activity classes using distance and time features.

To evaluate the impact of class balancing and feature selection on model performance, a scatter plot was applied to visualize the distribution of predicted activity classes based on transfer distance and time. Before applying the model figure 2.5, a significant overlap is observed between the 'low' and'medium' classes, indicating limited class separability despite a visible linear trend. Then after training the model using Gradient Boosting combined with the SMOTE Figure 2.6, a clearer clustering pattern emerges. The ‘High’ class is clearly located at distances greater than 2400 and times greater than 30, while the ‘Low’ and ‘Medium’ classes are mainly concentrated in the distance range 1000-1750 and the time interval 10-25. This improvement indicates that the model is able to distinguish patterns between classes more effectively after data balancing, resulting in more accurate and interpretative predictions based on relevant operational features.

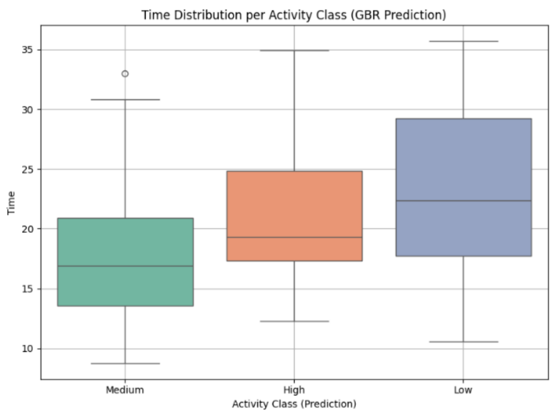

Figure 2. 7 Distribution of predicted transfer time across activity classes.

To further assess how the model distinguishes activity classes in terms of time distribution, a boxplot was generated for each predicted class, as shown in Figure 2.7. The visual indicates that the 'Low' class exhibits the widest spread, with a time range approximately from 11 to 34 and a higher interquartile variability. In contrast, the medium and high classes appear more compact, with the medium centered around a median of 18 and the high around

20. These results suggest that while the model has learned to assign transfer time ranges distinctly per class, the 'Low' class remains the most heterogeneous, possibly due to overlapping feature patterns with other categories. The overall separation of distributions supports the classifier's effectiveness in modeling transfer time variation across activity classes.

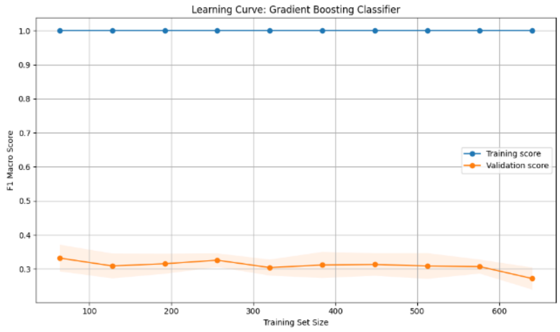

Figure 2. 8 Learning Curve before applying SMOTE and tuning.

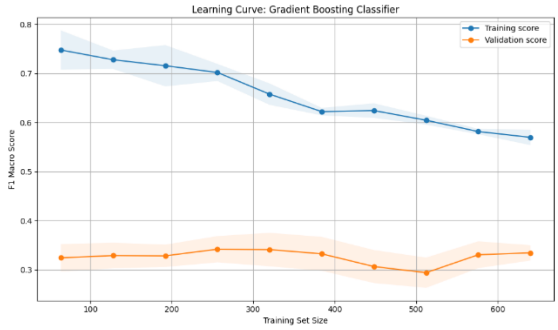

Figure 2. 9 Learning Curve after applying SMOTE and tuning..

The model's learning performance was then evaluated by comparing two learning curves before and after the application of SMOTE and hyperparameter optimization. As illustrated in Figure 2.8, before SMOTE, the model exhibited obvious overfitting, achieving a perfect F1 macro training score of 1.0 for all sizes of the training set, while the validation score remained low and unstable, fluctuating between 0.27 and 0.34. This gap reflects the limited generalizability of the model when trained on imbalanced activity class data. After applying SMOTE and tuning Figure 2.9, the training score showed a more realistic trend, starting at 0.75 and gradually decreasing to 0.57 as the data set grew. In particular, the validation score became more stable with a slight improvement, ranging between 0.29 and 0.35. The reduced disparity between training and validation performance indicates that the model was able to learn more generalizable patterns, validating the effectiveness of the optimization strategy in enhancing predictive accuracy for item transfer classification in activity-space-based scenarios.

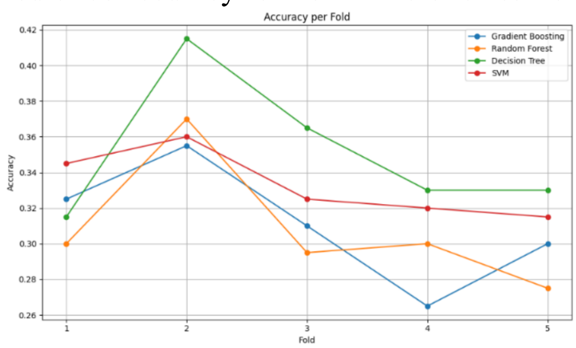

Figure 2. 10 Accuracy per fold DT, SVM, RF, Gradient Boosting

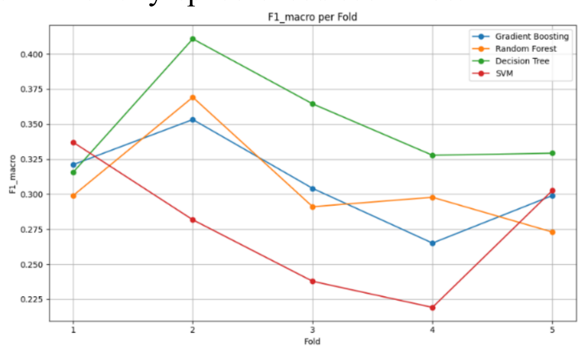

Figure 2. 11 F1_macro per fold DT, SVM, RF, Gradient Boosting

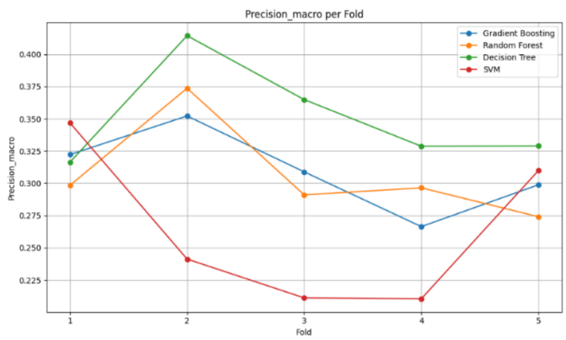

Figure 2. 12 Precision per fold DT, SVM, RF, Gradient Boosting

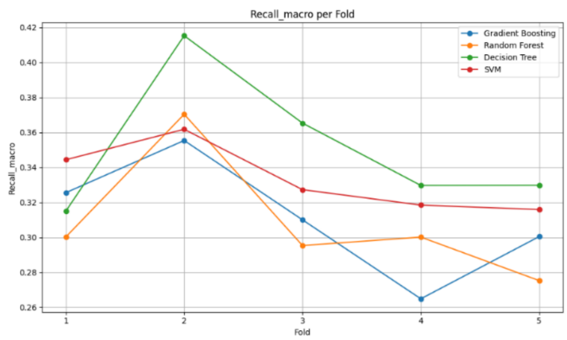

Figure 2. 13 Recall per fold DT, SVM, RF, Gradient Boosting

In addition to learning curve analysis, model comparisons were conducted across folds using four key metrics: Accuracy, Macro Precision, Macro Recall, and Macro F1 (Figures 2.10–2.13) to evaluate the performance of Gradient Boosting Classifier, Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM). Decision Tree consistently outperformed other models, achieving the highest accuracy (0.417), macro F1 (0.412), precision (0.412), and recall (0.417) in Fold 2, while maintaining stable results across all folds. Random Forest and Gradient Boosting showed moderate yet fluctuating performance, with their best results also occurring in Fold 2 (accuracy: 0.370 and 0.355; macro F1: 0.368 and 0.353; recall: 0.370 and 0.356). In contrast, SVM demonstrated the weakest and most inconsistent performance, particularly in macro F1 and precision, dropping to 0.219 and 0.21 respectively in Folds 3 and 4, highlighting its sensitivity to class imbalance. These findings confirm the robustness of tree-based models, particularly Decision Tree, in handling multi-class imbalanced datasets within activity-space logistics scenarios.

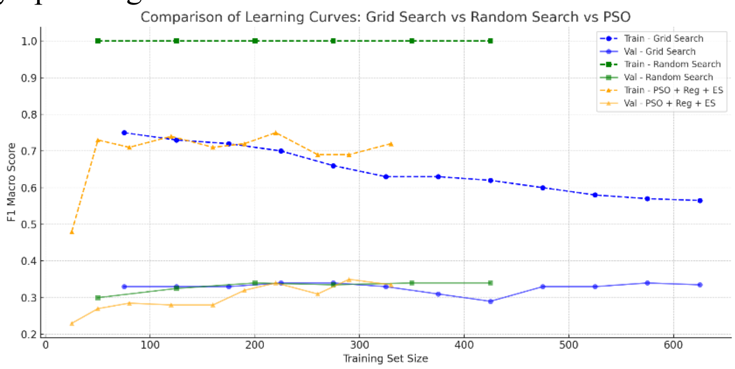

Figure 2. 14 PSO outperforms Grid and Random Search in validation performance

After comparing the performance of classification models, the next evaluation focused on the effectiveness of various hyperparameter tuning strategies using the learning curve shown in Figure 2.14. The results indicate that the PSO approach, combined with regularization and early stopping, yields the most optimal and stable performance, achieving the highest macro validation F1 score of approximately 0.74 and maintaining a consistent training score between 0.69 and 0.75 across all training set sizes. In contrast, Grid Search exhibits clear signs of overfitting, with a constant training score of 1.00 and a progressively decreasing validation score down to around 0.33 as the data size increases. Meanwhile, Random Search records the weakest performance, with both training and validation F1 scores stagnating at approximately 0.34 and 0.30, respectively. These findings underscore the superiority of PSO in generating models that are more generalizable and robust against increasing data complexity compared to conventional hyperparameter tuning methods. To assess the model's classification performance across all activity classes, confusion matrices were used to compare actual versus predicted labels before and after applying the Gradient Boosting model. This evaluation provides insights into the model's accuracy in classifying each category and helps identify areas where misclassifications are concentrated.

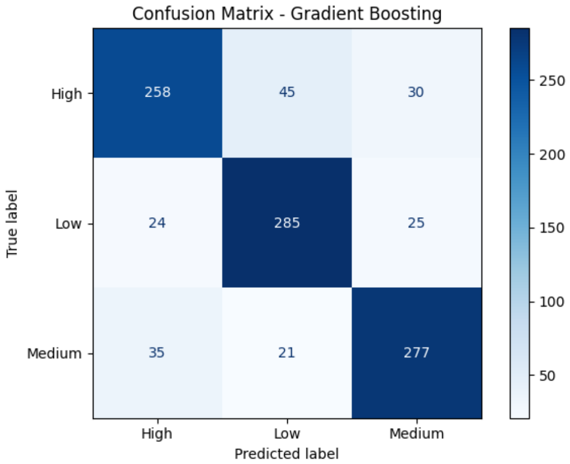

Figure 2. 15 Confusion matrix before applying the Gradient Boosting model.

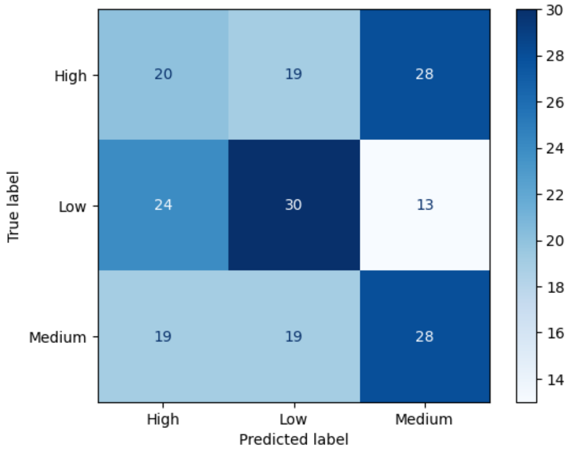

Figure 2. 16 Confusion matrix after Gradient Boosting model optimization

Before model application Figure 2.15, class predictions exhibited high confusion, particularly between 'high' and 'medium', as well as between'medium' and 'low.' For example, only 20 out of 67 'high' instances were correctly predicted, while a significant number were misclassified as ‘Medium’. Similar patterns are seen in the 'Low' and 'Medium' classes, suggesting that the model struggled to establish clear boundaries.

After applying the optimized gradient boost model with SMOTE Figure 2.16, classification accuracy improved substantially. The number of correct predictions increased for all classes, with the 'high', 'low', and'medium' classes achieving 258, 285, and 277 correct predictions, respectively. Misclassifications decreased considerably, especially for the'medium', which had previously been the most ambiguous. This confirms that the model, after tuning and balancing, was able to better capture class-specific patterns and minimize overlap.

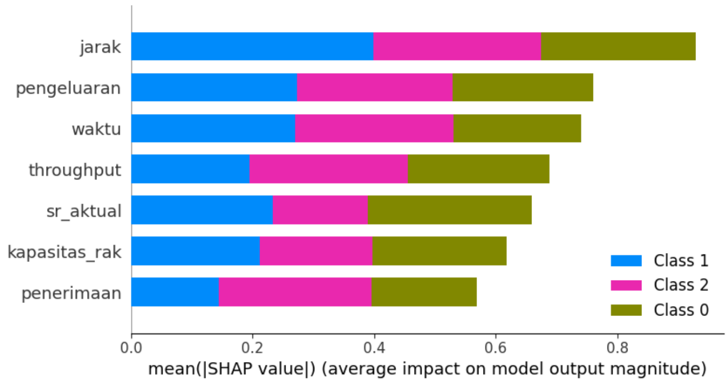

Figure 2. 17 SHAP summary plot showing feature importance across predicted activity classes

To improve the interpretability of the Gradient Boosting model, SHAP was used to measure the average contribution of each feature to the prediction output. The SHAP summary plot in Figure 2.17 shows that distance is the most influential feature with a mean SHAP value close to 0.9, followed by distribution volume and transfer time, each contributing approximately 0.75. Other features such as throughput, actual service rate, and rack capacity each contribute between 0.65 to 0.70, while receiving shows the lowest influence at around 0.55. These results confirm that spatial and operational variables, particularly distance, play a central role in distinguishing activity classes, supporting earlier findings from both correlation and classification performance evaluations. By identifying key drivers such as throughput, rack capacity, and actual service level, the model provides actionable insights for warehouse managers to optimize spatial layout, reduce lead times, and improve SLA (Service Level Agreement) compliance across high-volume distribution networks.

DISCUSSION

Although the optimized gradient boost model showed strong performance in classifying the transfer time of the elements, several limitations were identified. In particular, class overlap, especially between 'low' and 'medium', persisted despite SMOTE and tuning, indicating that the current set of characteristics may lack contextual depth. Incorporating additional variables such as operator workload, congestion, or temporal patterns (e.g., shift or weekday) could improve class separability. In practical terms, the predictive classification of item transfer time enables warehouse management to proactively adjust storage zones and dynamically schedule picking activities. For instance, goods with a 'High' predicted transfer time can be assigned to locations closer to dispatch areas to reduce handling delays.

The model's reliance on numerical inputs also raises concerns for real-world deployment, where sensor data may be incomplete or delayed. Integrating categorical or time-based features may improve robustness. Moreover, while SMOTE addressed imbalance during training, its synthetic samples may not fully capture rare class variability; therefore, alternative methods such as cost-sensitive learning could be explored.

Tabel 1. Compared with related works summarized

Author	Model	Preprocessing	SMOTE	Feature Importance	Hyperparameter Tuning
(Lin et al., 2023)	Polynomial, Degree 3	Sample size filtering, normalization	X	X	X
(Tufano et al., 2022)	Random Forest	Handling missing value	X	✓	✓
(Khiari & Olaverri-Monreal, 2020)	Histogram-Based Gradient Boosting	Feature Extraction, Outliers removal	X	X	X
(Ye et al., 2024)	Spatio-Temporal Graph Convolutional Network	Normalization, Outliers removal	X	X	X
This Study	Gradient Boosting	Data Normalization, Standarization, Pearson Corelation	✓	✓	✓

Compared with related works summarized in Table I, each study exhibits a different focus on modeling. Compared with related works summarized in Table I, each study exhibits a different focus on modeling. Lin et al. (Lin et al., 2023) used a polynomial regression model with basic preprocessing approaches such as normalization and sample size filtering, but did not include hyperparameter adjustment or feature importance analysis. The approach of Tufano et al. (Tufano et al., 2022) is more interpretative, utilizing Random Forest with grid search and Permutation Importance. Khiari et al. (Khiari & Olaverri-Monreal, 2020) applied Histogram-Based Gradient Boosting with a primary focus on robust regression through outlier removal, without explicitly adding tuning or feature exploration. Ramanujam et al. (Ye et al., 2024) adopt a Graph Neural Network (ST-GCN)-based approach to model dynamic patterns in travel time prediction, focusing on spatio-temporal structure and outlier detection. Unlike commonly used regression approaches, the method proposed in this study shifts to Gradient Boosting-based classification, integrating SMOTE for class balancing, SHAP for feature interpretability, and grid search for hyperparameter tuning. This approach reflects methodological advances focused on accuracy, fairness of class distribution, and transparency in item transfer classification models.

CONCLUSION

This study shows that combining normalization and SMOTE with an optimized Gradient Boosting model effectively classifies the item transfer time using spatial and operational features. Distance was the most impactful variable (SHAP ≈ 0.90), followed by distribution volume (0.75), transfer time (0.74), and throughput (0.70), confirming their importance in prediction.

After adjustment, the model correctly classified 258 'high', 285 'low', and 277 'medium' instances, significantly better than the baseline, where only 20 'high' cases were correctly predicted. Visual tools like scatter plots and box plots confirmed better class separation, especially for ‘High’ at distances greater than 2400 and times over 30. Despite improvements, some confusion persisted between 'low' and 'medium', indicating the need for additional contextual features. SMOTE improved balance, but may not fully represent the variability of real-world data.

In conclusion, the study presents a structured and explainable framework for predicting transfer time in warehouse settings. Future work may also expand beyond classification by combining it with regression or time series modeling to estimate transfer time more precisely, particularly for borderline cases. These enhancements would improve model accuracy and practical relevance in operational decision-making. From a managerial perspective, this predictive modeling approach supports informed decision-making in resource planning, labor allocation, and layout restructuring. It aligns with the goals of modern warehouse management systems to increase agility, responsiveness, and efficiency in item handling and fulfillment processes.

REFERENCE

Bahmani, M., Shawi, R. E., Potikyan, N., & Sakr, S. (2021). *To tune or not to tune? an approach for recommending important hyperparameters*. https://doi.org/10.48550/arxiv.2108.13066

Bischl, B., Binder, M., Lang, M., Pielok, T., Richter, J., Coors, S., Thomas, J., Ullmann, T., Becker, M., Boulesteix, A., Deng, D., & Lindauer, M. (2023). Hyperparameter optimization: foundations, algorithms, best practices, and open challenges. *WIREs Data Mining and Knowledge Discovery*, *13*(2). https://doi.org/10.1002/widm.1484

Boldini, D., Friedrich, L., Kuhn, D., & Sieber, S. A. (2022). Tuning gradient boosting for imbalanced bioassay modelling with custom loss functions. *Journal of Cheminformatics*, *14*(1), 80. https://doi.org/10.1186/s13321-022-00657-w

Brijith, A. (2023). *Data Preprocessing for Machine Learning*. 2023.

de Amorim, L. B. V, Cavalcanti, G. D. C., & Cruz, R. M. O. (2022). The choice of scaling technique matters for classification performance. *ArXiv Preprint ArXiv:2212.12343*. https://arxiv.org/abs/2212.12343

Derby, H., Chander, H., Kodithuwakku Arachchige, S. N. K., Turner, A. J., Knight, A. C., Burch, R., Freeman, C., Wade, C., & Garner, J. C. (2023). Occupational Footwear Design Influences Biomechanics and Physiology of Human Postural Control and Fall Risk. *Applied Sciences*, *13*(1). https://doi.org/10.3390/app13010116

Fan, S., Guerlet, S., Forget, F., Bierjon, A., Millour, E., Ignatiev, N., Shakun, A., Grigoriev, A., Trokhimovskiy, A., Montmessin, F., & Korablev, O. (2022). Thermal Tides in the Martian Atmosphere Near Northern Summer Solstice Observed by ACS/TIRVIM Onboard TGO. *Geophysical Research Letters*, *49*(7). https://doi.org/10.1029/2021gl097130

Farhadpour, S., Warner, T. A., & Maxwell, A. E. (2024). Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices. *Remote Sensing*, *16*(3), 533. https://doi.org/10.3390/rs16030533

Ilemobayo, J. A., Durodola, O. I., Alade, O., Awotunde, O. J., Olanrewaju, A. T., Falana, O., Ogungbire, A., Osinuga, A., Ogunbiyi, D., Ifeanyi, A., Odezuligbo, I. E., & Edu, O. E. (2024). Hyperparameter tuning in machine learning: a comprehensive review. *Journal of Engineering Research and Reports*, *26*(6), 388–395. https://doi.org/10.9734/jerr/2024/v26i61188

Iqbal, S., Qureshi, A. N., Ullah, A., Li, J., & Mahmood, T. (2022). Improving the robustness and quality of biomedical cnn models through adaptive hyperparameter tuning. *Applied Sciences*, *12*(22), 11870. https://doi.org/10.3390/app122211870

Khiari, J., & Olaverri-Monreal, C. (2020). Boosting Algorithms for Delivery Time Prediction in Transportation Logistics. *2020 International Conference on Data Mining Workshops (ICDMW)*, 251–258. https://doi.org/10.1109/ICDMW51313.2020.00043

Kocsis Szürke, S., Sütheö, G., Apagyi, A., Lakatos, I., & Fischer, S. (2022). Cell Fault Identification and Localization Procedure for Lithium-Ion Battery System of Electric Vehicles Based on Real Measurement Data. *Algorithms*, *15*(12). https://doi.org/10.3390/a15120467

Kumar, A., & Singh, R. (2023). Interpretable Machine Learning Using SHAP: Recent Advances and Applications. *Information Sciences*, *639*, 119–134.

Lin, Y.-K., Chen, C.-F., & Chou, T.-Y. (2023). Developing Prediction Model of Travel Times of the Logistics Fleets of Large Convenience Store Chains Using Machine Learning. *Algorithms*, *16*(6), 286.

Liu, Y., & Chen, H. (2022). An Improved SMOTE-Tomek Links Algorithm for Imbalanced Data Classification. *Expert Systems with Applications*, *188*, 115963.

Lundberg, S. ~M., & Lee, S.-I. (2020). A unified approach to interpreting model predictions. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-1

Mohapatra, S., Sasy, S., He, X., Kamath, G., & Thakkar, O. (2021). *The role of adaptive optimizers for honest private hyperparameter selection*. https://doi.org/10.48550/arxiv.2111.04906

Opitz, J. (2024). A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice. *Transactions of the Association for Computational Linguistics*. https://doi.org/10.1162/tacl_a_00675

Pane, S. F., Putrada, A. G., Alamsyah, N., & Fauzan, M. N. (2022). A PSO-GBR solution for association rule optimization on supermarket sales. *2022 Seventh International Conference on Informatics and Computing (ICIC)*, 1–6.

Ramadhan, B., & Pane, S. F. (2024). Pengaruh Hyperparameter Tuning untuk Efektivitas pada Pendekatan Hybrid dalam Mendiagnosis Stres dan Depresi: Tinjauan Studi Literatur. *Jurnal Tekno Insentif*, *18*(2), 104–118.

Saito, T., & Rehmsmeier, M. (2020). The precision–recall plot is more informative than the ROC plot when evaluating classifiers on imbalanced datasets. *PLOS ONE*, *15*(3), e0233147. https://doi.org/10.1371/journal.pone.0233147

Salih, A. M., Raisi-Estabragh, Z., Galazzo, I. B., Radeva, P., Petersen, S. E., Lekadir, K., & Menegaz, G. (2024). A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *Advanced Intelligent Systems*, *7*(1). https://doi.org/10.1002/aisy.202400304

Song, Y., Gu, Y., Guo, H., Yang, H., Wang, X., Wu, H., Wang, A., Wang, H., Zhang, Q., Zhang, Q., Liu, L., Meng, G., Liu, B., & Niu, K. (2023). Association Between Neutrophil-to-Lymphocyte Ratio and Benign Prostatic Hyperplasia: Results from the TCLSIH Cohort Study. *Journal of Inflammation Research*, *16*, 4857–4866. https://doi.org/10.2147/JIR.S431049

Su, N., Huang, S., & Su, C. (2024). Elevating Smart Manufacturing with a Unified Predictive Maintenance Platform: The Synergy between Data Warehousing, Apache Spark, and Machine Learning. *Sensors*, *24*(13). https://doi.org/10.3390/s24134237

Tang, B., Pan, Z., & Yin, K. (2023). Explainable Machine Learning for Healthcare: A Case Study with SHAP. *Applied Soft Computing*, *140*, 110287. https://doi.org/10.1016/j.asoc.2023.110287

Tufano, A., Accorsi, R., & Manzini, R. (2022). A machine learning approach for predictive warehouse design. *The International Journal of Advanced Manufacturing Technology*, *119*(3), 2369–2392.

Xiong, Y., Liu, X., Lan, L., You, Y., Si, S., & Hsieh, C. (2020). *How much progress have we made in neural network training? a new evaluation protocol for benchmarking optimizers*. https://doi.org/10.48550/arxiv.2010.09889

Ye, T., Cheng, S., Hijazi, A., & Van Hentenryck, P. (2024). Contextual Stochastic Optimization for Omnichannel Multi-Courier Order Fulfillment Under Delivery Time Uncertainty. *ArXiv Preprint ArXiv:2409.06918*.

Zhang, W., & Li, J. (2023). Enhanced Particle Swarm Optimization for Hyperparameter Tuning in Machine Learning Models. *Applied Soft Computing*, *136*, 110012.

Zhang, Y., Deng, L., & Wei, B. (2024). Imbalanced Data Classification Based on Improved Random-SMOTE and Feature Standard Deviation. *Mathematics*, *12*(11). https://doi.org/10.3390/math12111709