JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)
VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8282

PSO-OPTIMIZED XGBOOST FOR MAIL DELIVERY DELAY PREDICTION AND LOGISTICS SLA COMPLIANCE

Syafrial Fachrie Pane[1*]; Bargana Kukuh Raditya[2]

School of Information Technology, Applied Bachelor Program of Informatics Engineering[1,2]
Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia[1,2]
https://ulbi.ac.id/[1,2]
syafrial.fachri@ulbi.ac.id*, barganakukuh@gmail.com

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *The Indonesian postal logistics sector continues to face challenges in maintaining letter delivery timeliness and Service Level Agreement (SLA) compliance under dynamic operational conditions. Conventional predictive approaches often rely on static representations and are limited in capturing temporal risk patterns in mail delivery processes. This study proposes a data-driven machine learning framework to predict delivery delay risk in postal letter services by integrating temporal feature engineering, class imbalance handling, and metaheuristic-based hyperparameter optimization. The framework applies the Synthetic Minority Over-sampling Technique (SMOTE) and evaluates multiple classification models using stratified cross-validation. Among the evaluated algorithms, XGBoost optimized using Particle Swarm Optimization (PSO) demonstrates the strongest predictive performance. The PSO-optimized configuration (n = 96, lr = 0.1718, d = 3) achieves an accuracy of 0.6605, ROC–AUC of 0.6883, and F1-score of 0.6059, indicating improved class-sensitive prediction. Model interpretability is examined using Mean Decrease in Impurity (MDI), which identifies posting day as the dominant contributor to delivery delays, followed by SLA commitment and intra-day posting patterns. The final framework generates probabilistic risk scores from 0 to 100 percent, with the highest observed value reaching 99.44, enabling early warning and prescriptive operational interventions for potential SLA violations. These results indicate that the proposed PSO–XGBoost framework supports proactive logistics risk management. However, this study is limited to historical data from a single postal operational environment and does not incorporate external factors such as weather, traffic, or regional delivery variations.*

Keywords: *Delivery Delay Risk Prediction, Particle Swarm Optimization, Postal Letter Logistic, Service Level Agreement (SLA), XGBoost.*

Intisari—*Sektor logistik pos di Indonesia masih menghadapi tantangan dalam menjaga ketepatan waktu pengiriman surat dan kepatuhan terhadap Service Level Agreement (SLA) di bawah kondisi operasional yang dinamis. Pendekatan prediktif konvensional umumnya bergantung pada representasi statis dan masih terbatas dalam menangkap pola risiko temporal pada proses pengiriman surat. Penelitian ini mengusulkan kerangka kerja machine learning berbasis data untuk memprediksi risiko keterlambatan pengiriman pada layanan surat pos dengan mengintegrasikan rekayasa fitur temporal, penanganan ketidakseimbangan kelas, dan optimasi hiperparameter berbasis metaheuristik. Kerangka kerja ini menerapkan Synthetic Minority Over-sampling Technique (SMOTE) dan mengevaluasi beberapa model klasifikasi menggunakan stratified cross-validation. Di antara algoritma yang dievaluasi, XGBoost yang dioptimasi menggunakan Particle Swarm Optimization (PSO) menunjukkan kinerja prediktif terbaik. Konfigurasi hasil optimasi PSO (n = 96, lr = 0,1718, d = 3) mencapai akurasi sebesar 0,6605, ROC–AUC sebesar 0,6883, dan F1-score sebesar 0,6059, yang menunjukkan peningkatan prediksi yang sensitif terhadap kelas. Interpretabilitas model dianalisis menggunakan Mean Decrease in Impurity (MDI), yang mengidentifikasi hari pengiriman sebagai kontributor dominan terhadap keterlambatan, diikuti oleh komitmen SLA dan pola pengiriman dalam satu hari. Kerangka kerja akhir menghasilkan skor risiko probabilistik dalam rentang 0 hingga 100 persen, dengan nilai tertinggi* *mencapai 99,44, sehingga mendukung peringatan dini dan intervensi operasional terhadap potensi pelanggaran SLA. Hasil penelitian menunjukkan bahwa kerangka kerja PSO–XGBoost mendukung manajemen risiko logistik yang proaktif. Namun, penelitian ini masih terbatas pada data historis dari satu lingkungan operasional pos dan belum mengintegrasikan faktor eksternal seperti cuaca, lalu lintas, atau variasi wilayah pengiriman.*

Kata Kunci*: Logistik Surat Pos, Optimasi Kawanan Partikel, Prediksi Risiko Keterlambatan Pengiriman, Tingkat Layanan (SLA), XGBoost.*

INTRODUCTION

Since the rapid growth of postal services and document-based delivery, timely letter delivery has become a critical factor in ensuring service reliability and operational efficiency. In national postal systems, delays in letter delivery directly impact public trust and compliance with predefined Service Level Agreements (SLA) [1], [2]. The increasing volume of mailed documents and the operational complexity of postal delivery networks have made letter distribution systems more vulnerable to delays, particularly under dynamic operational conditions. Similar to complex phenomena observed during large-scale service systems, postal delivery performance is influenced by multiple interrelated temporal, operational, and service-level factors, making accurate delay prediction a challenging task. Previous studies have shown that traditional statistical and rule-based approaches often struggle to capture non-linear relationships and complex interactions within large-scale postal operational data. Conventional prediction models frequently fail when dealing with heterogeneous features and dynamic posting behavior, necessitating the adoption of more robust machine learning approaches capable of handling such complexity.

In postal letter services, delivery delay is not solely determined by service duration but is also strongly influenced by posting time, posting day, weekend effects, and service commitment categories. These factors interact in non-trivial ways and play a crucial role in determining SLA compliance [3]. Failure to account for such temporal and operational dependencies often leads to inaccurate delay predictions and ineffective SLA management. Ensemble machine learning methods have emerged as powerful tools for predictive modeling in complex service domains [4]. Methods including Random Forest, Gradient Boosting, and XGBoost are widely recognized for their strong predictive capability, particularly in capturing non-linear relationships and complex feature dependencies. Among these, XGBoost has gained particular attention for postal and logistics applications because of its scalability, built-in regularization mechanisms, and strong generalization capability [5], [6]. However, the effectiveness of ensemble models remains highly dependent on optimal hyperparameter configurations, which are often difficult to obtain through manual tuning or exhaustive search. To address this challenge, metaheuristic based optimization approaches, including Particle Swarm Optimization (PSO), have been widely adopted. PSO offers an efficient mechanism for exploring high-dimensional hyperparameter spaces by simulating collective behavior, allowing ensemble models to converge toward optimal configurations [7]. Prior research confirms that integrating PSO with ensemble learning significantly improves predictive accuracy and robustness [8]. In the context of postal letter delivery, such optimization is essential for reliably identifying potential SLA violations.

Beyond predictive accuracy, understanding the root causes of letter delivery delays is equally important for postal operational decision-making. Feature importance analysis provides insights into which variables most strongly influence model predictions [9], [10]. Techniques such as Mean Decrease in Impurity (MDI) enable the interpretation of ensemble models and the identification of dominant factors contributing to postal SLA non-compliance [11]. This interpretability supports actionable interventions, such as adjusting posting schedules or refining service-level commitments. Despite the extensive application of ensemble learning and optimization techniques in various domains, their combined use for postal SLA violation prediction and root cause analysis remains underexplored, particularly in developing postal logistics environments [12]. Most existing studies focus primarily on improving prediction accuracy without systematically examining feature contributions. This research addresses this gap by developing a PSO-optimized XGBoost model specifically designed to predict letter delivery SLA violations and to identify key drivers of postal delivery delays.

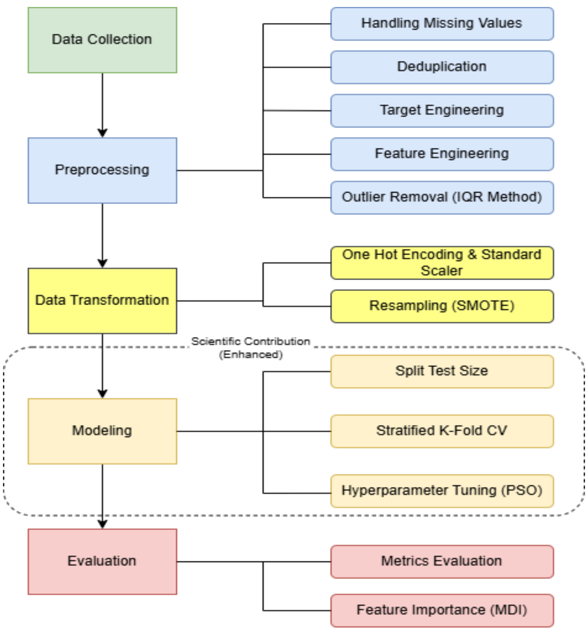
This study makes three key contributions. First, it develops an end-to-end machine learning pipeline integrating data cleaning, P50-based SLA target engineering, IQR-based outlier removal, and

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8282

SMOTE-based class balancing for postal letter delivery data. Second, it conducts a comparative evaluation of six classification models under stratified K-Fold CV with Grid Search, Random Search, and PSO-based hyperparameter tuning. Third, it analyzes feature importance using the best-performing PSO-optimized XGBoost model to identify dominant drivers of postal SLA violations. The proposed framework generates binary delivery status predictions (On-Time or OverSLA) along with calibrated risk probabilities (0–100%), enabling early risk detection and proactive SLA management in postal operations.

MATERIALS AND METHODS

Figure 1 presents the methodological framework employed in this study, which consists of five main stages: (1) data collection, (2) data preprocessing, (3) data transformation, (4) modeling and hyperparameter optimization, and (5) evaluation and interpretation. The dependent variable is is_oversla, a binary indicator representing logistics SLA compliance (on-time vs. delayed). The preprocessing stage includes missing value handling, deduplication, target engineering, feature extraction, and outlier removal using the Interquartile Range (IQR) method. Data transformation is performed through One-Hot Encoding for categorical variables, feature standardization for numerical attributes, and class balancing using the Synthetic Minority Over-sampling Technique (SMOTE).



Source: (Research Results, 2025)
Figure 1. Methodological Framework of the Study

1. Data Collection.

This study employs a dataset obtained from a major logistics service provider in Indonesia, focusing on mail delivery transactions handled by the Subdirectorate of Last Mile Operations at the Bandung Head Office. The dataset comprises 3,000 raw transaction records collected between August 1, 2025, and October 30, 2025. For confidentiality purposes, shipment tracking numbers are used exclusively as anonymized identifiers. The model inputs include service type and temporal features, namely posting hour, posting day, and a weekend indicator encoded as a binary variable. Product-specific reference SLA values are incorporated as dynamic baseline features derived from historical process-level performance. The target variable, is_oversla, is defined as a binary indicator of SLA violation, where deliveries exceeding the product-specific reference SLA are labeled as High Risk (1), and those within the threshold are labeled as On-Time (0).

2. Data Preprocessing.

During the preprocessing stage, this study applies a set of methods to construct an adaptive target variable, control the influence of extreme observations, and extract relevant temporal features to support SLA violation prediction. This stage is focused on the development of data-driven SLA thresholds using a P50 (median-based), binary target construction, outlier removal, and calendar-based feature engineering. A binary target variable y is defined to represent the delivery risk status based on SLA compliance. The target formulation follows a failure-based classification logic, where SLA violations are treated as high-risk events [13]. A binary target variable y is defined to represent the delivery risk status based on SLA compliance.

$$\mathrm{SLA}_{\mathrm{p}}^{\mathrm{ref}} = \mathrm{median}\left(\mathrm{SLA}_{\mathrm{i,p}}^{\mathrm{pickup}}\right) + \mathrm{median}\left(\mathrm{SLA}_{\mathrm{i,p}}^{\mathrm{linehaul}}\right) \tag{1}$$

The reference SLA value in Equation 1 is defined using a P50 (median based) aggregation of the pickup, linehaul, and last-mile SLA components for each product or service type p, where the index i enumerates individual delivery samples within the same product category rather than representing temporal order. The resulting $\mathrm{SLA}_{\mathrm{p}}^{\mathrm{ref}}$ establishes a robust decision threshold for binary SLA labeling, where deliveries exceeding the threshold are classified as High Risk, and those within the threshold are classified as On-Time. Based on the reference SLA threshold defined in Equation 1, the

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8282

binary classification target is constructed as formalized in Equation 2.

$$\mathrm{is\_oversla_i} = \begin{cases} 1, \mathrm{SLA}_i^{\mathrm{actual}} > \mathrm{SLA}_p^{\mathrm{ref}} \\ 0, \mathrm{SLA}_i^{\mathrm{actual}} \leq \mathrm{SLA}_p^{\mathrm{ref}} \end{cases} \quad (2)$$

Equation 2 assigns a binary risk label to each delivery instance, where $(y = 1)$ indicates a delivery exceeding the product-specific reference SLA (High Risk), and $(y = 0)$ denotes an on-time delivery. In Equation 3, the Interquartile Range (IQR) is defined as a statistical method is applied to detect and remove extreme delivery delays [14]. Variability in delivery duration is quantified through the interquartile range, which captures the spread between Q_1 and Q_3.

$$\mathrm{IQR} = \mathrm{Q}_3 - \mathrm{Q}_1 \quad (3)$$

The integration of adaptive SLA-based target construction, IQR-based outlier removal, and temporal feature engineering results in a clean and structured dataset that is suitable for robust predictive modeling and reliable SLA violation analysis.

3. Classification Modeling with SMOTE and Evaluation

A set of six classification algorithms, namely Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Extra Trees (ET), Gradient Boosting (GB), and XGBoost, is evaluated to model non-linear operational risk patterns in logistics delivery data. To mitigate class imbalance between delayed and on-time shipments, The SMOTE procedure is restricted to the training data of each fold to mitigate class imbalance [15]. Model performance is assessed using Stratified K-Fold CV, which preserves class proportions across all folds to ensure unbiased and stable performance estimation.

Among the evaluated models, XGBoost is selected as the primary classification architecture due to its superior randomization and regularization mechanisms, which effectively reduce variance and enhance robustness against overfitting [5]. Unlike conventional ensemble methods, XGBoost constructs an additive model by sequentially combining multiple decision trees optimized via gradient-based learning [6]. The prediction function of XGBoost is defined in Equation 4:

$$\hat{y}_i = \sum_{m=1}^{M} \mathrm{f_m}(\mathrm{x_i}), \quad \mathrm{f_m} \in \mathcal{F} \quad (4)$$

In Equation 4, $\hat{y}_i$ denotes the predicted output for instance i, M represents the number of boosting trees, and each $\mathrm{f_m}$ corresponds to a regression tree belonging to the functional space $\mathcal{F}$. Model training is guided by a regularized objective function that balances empirical loss and model complexity, as formalized in Equation 5:

$$\mathcal{L} = \sum_{i=1}^{N} \mathrm{l}(\mathrm{y_i}, \hat{y}_i) + \sum_{m=1}^{M} \Omega(\mathrm{f_m}) \quad (5)$$

Here, l denotes the loss function measuring prediction error, while $\Omega(\mathrm{f_m})$ is a regularization term that penalizes tree complexity through structural constraints such as tree depth and leaf weights [6]. This formulation enables XGBoost to capture complex non-linear relationships while maintaining generalization performance under imbalanced logistics data.

By minimizing the regularized loss function and impurity-based split criteria, the model maximizes information gain, thereby enabling effective discrimination between On-Time and High-Risk shipments. Final model performance is validated using Accuracy, Macro-Average F1-Score, and ROC–AUC to ensure reliable predictive capability and robust classification quality [16].

4. Hyperparameter Optimization via Particle Swarm Optimization

Hyperparameter optimization is conducted through a three-phase strategy consisting of Grid Search, Random Search, and PSO [5], [17], [18]. The particle velocity and position updates are defined in Equations 6 and 7, which jointly govern the search dynamics within the hyperparameter space:

$$\mathrm{v_i}(\mathrm{t}+1) = \mathrm{w} \cdot \mathrm{v_i}(\mathrm{t}) + \mathrm{c_1 r_1}\left(\mathrm{p_{best,i}} - \mathrm{x_i}(\mathrm{t})\right) + c\_2\, r\_2\, (g\_best - x\_i\ (t)) \quad (6)$$

$$\mathrm{x_i}(\mathrm{t}+1) = \mathrm{x_i}(\mathrm{t}) + \mathrm{v_i}(\mathrm{t}+1) \quad (7)$$

In these equations, $\mathrm{v_i}(\mathrm{t})$ and $\mathrm{x_i}(\mathrm{t})$ correspond to the motion parameters of particle i, while $\mathrm{p_{best,i}}$ and $\mathrm{g_{best}}$ denote the best solutions found at the individual swarm levels. The parameter w governs convergence behavior by balancing exploration and exploitation, whereas stochastic coefficients $\mathrm{r_1}$ and $\mathrm{r_2}$ sampled from the unit interval, enhance the algorithm's global search capability.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8282

Through iterative updates of Equations 6 and 7, the PSO algorithm converges toward optimal values of *n_estimators*, *max_depth*, and *learning_rate* by maximizing the F1-score under stratified cross-validation.

5. Feature Importance Analysis: Mean Decrease Impurity (MDI)

Feature importance is quantified using MDI criterion, which evaluates the contribution of each feature to impurity reduction in tree-based models [19], [20]. The MDI formulation is defined in Equation 8:

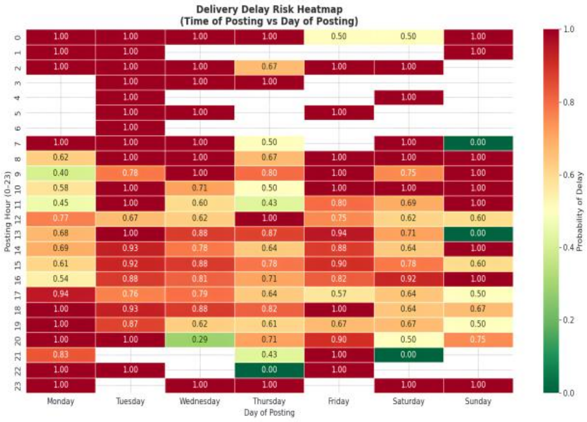
$$\mathrm{MDI}(k,\mathcal{T}) = \sum_{t\in\mathcal{I}(\mathcal{T}),s(t)=k} \frac{n(t)}{n} \Delta\mathcal{I}(t) \quad (8)$$

In Equation 8, the summation aggregates impurity reductions over all internal nodes $t \in \mathcal{I}(\mathcal{T})$ where feature k is selected as the split variable. The weighting term $n(t)/n$ indicates the relative sample weight at node t, while $\Delta\mathcal{I}(t)$ captures the magnitude of impurity reduction produced by that split. This formulation yields a normalized measure of feature relevance within the decision tree.

RESULTS AND DISCUSSION

In the initial stage of the analysis, data preparation and target definition are performed to establish a reliable modeling foundation. From the original 3,000 mail delivery records, data cleaning removes duplicated entries and records with missing values, resulting in 2,000 valid observations. A subsequent SLA processing stage eliminates records with incomplete or logically inconsistent SLA information, reducing the dataset to 1,135 observations. Finally, extreme SLA values are filtered using the Interquartile Range (IQR) method, producing a final dataset of 1,074 observations with improved consistency and distributional stability. Following data cleaning, the binary target variable is_oversla is defined in accordance with Equation 1. Shipments exceeding the SLA threshold are labeled as high-risk deliveries ($y = 1$), while shipments completed within the SLA limit are classified as on-time ($y = 0$). This target formulation enables a clear separation between delayed and compliant deliveries and supports subsequent predictive analysis.

In the second stage, exploratory data analysis (EDA) is conducted to identify temporal patterns associated with mail delivery delays [21], [22]. Figure 2 presents a delivery delay risk heatmap based on the interaction between posting hour and posting day. The color intensity represents the probability of delay, where red indicates higher delay risk and green indicates lower delay risk.



Source: (Research Results, 2025)
Figure 2. Risk Heatmap: Delay Clusters in Late Posting Hours

The heatmap shows that high delay probabilities are concentrated during early morning hours, particularly between 00:00 and 02:00 on Monday to Wednesday. Elevated delay risks also appear from midday to evening, especially between 12:00 and 19:00 on Friday and Saturday. In contrast, lower delay risks are observed in several Thursday time windows, suggesting more stable delivery conditions on that day. Blank cells indicate unavailable observations for specific posting hour and day combinations rather than zero delay risk. These findings demonstrate that delay risk is strongly influenced by temporal posting patterns, thereby supporting the inclusion of posting hour and posting day as key predictive features in the modeling stage.

In the third stage, predictive modeling is conducted using six machine learning algorithms, namely Decision Tree, Extra Trees, Gradient Boosting, Logistic Regression, Random Forest, and XGBoost. Model robustness is evaluated using Stratified K-Fold CV across multiple test split ratios (0.1, 0.2, 0.3, and 0.4) to identify the most effective data partitioning strategy [23], [24]. Hyperparameter optimization is performed using Grid Search, Random Search, and PSO, while model performance is assessed based on Accuracy, ROC–AUC, and F1-Score [25]. As shown in Table 1, the comparative evaluation indicates that ensemble-based models generally outperform the linear baseline model in predicting mail delivery delay risk. XGBoost and Gradient Boosting produce the most competitive results across the evaluated optimization strategies, suggesting their stronger capability to capture non-linear temporal and operational patterns. Among all configurations, PSO-XGBoost provides the most balanced performance across Accuracy, ROC-AUC, and F1-Score, supporting its selection as the best-performing model in this study. In contrast, Logistic Regression consistently records the lowest performance, indicating limited suitability for modeling complex SLA violation patterns. These findings provide the basis for the more detailed XGBoost tuning analysis presented in Table 2.

Table 1 Performance Comparison of Classification Models under Different Hyperparameter Optimization Methods

Model	Method	Accuracy	ROC-AUC	F1-Score
Decision Tree	Grid Search	0.6232	0.6228	0.5694
Extra Trees	Grid Search	0.6232	0.6302	0.5789
Gradient Boosting	Grid Search	0.6372	0.6877	0.5955
Logistic Regression	Grid Search	0.5534	0.5825	0.5268
Random Forest	Grid Search	0.6232	0.6254	0.5661
XGBoost	Grid Search	0.6186	0.6853	0.5802
Decision Tree	Random Search	0.6186	0.6640	0.5802
Extra Trees	Random Search	0.6232	0.6302	0.5661
Gradient Boosting	Random Search	0.6511	0.6774	0.6013
Logistic Regression	Random Search	0.5534	0.5828	0.5268
Random Forest	Random Search	0.6232	0.6308	0.5661
XGBoost	Random Search	0.6372	0.6415	0.5838
Decision Tree	PSO	0.6232	0.6360	0.5694
Extra Trees	PSO	0.5813	0.6641	0.5520
Gradient Boosting	PSO	0.6465	0.6848	0.5945
Logistic Regression	PSO	0.5534	0.5827	0.5268
Random Forest	PSO	0.6232	0.6284	0.5661
XGBoost	**PSO**	**0.6605**	**0.6883**	**0.6059**

Source: (Research Results, 2025)

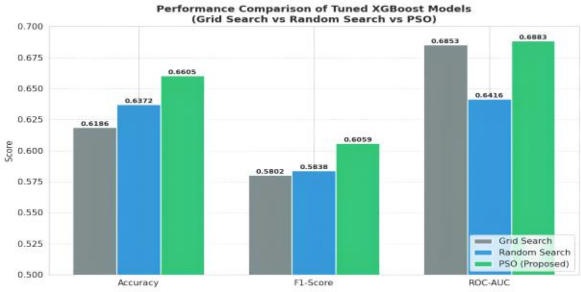
As summarized in Table 2, statistical analysis identifies the XGBoost model tuned via PSO as the most effective configuration for the dataset. As shown in the optimized model summary, the proposed PSO-XGBoost configuration achieved an Accuracy of 0.6605, an ROC-AUC of 0.6883, and an F1-Score of 0.6059. The optimal hyperparameter set was determined as *n*=96, *lr*=0.1718, and *d*=3, with a computational execution time of 21.3482 seconds. This configuration demonstrates the most effective balance between predictive precision and computational efficiency for the specified logistics environment.

Table 2 Performance Comparison of XGBoost Hyperparameter Tuning Methods

Method	Best Parameters	Accuracy	ROC-AUC	F1-Score	Time(s)
Grid Search	lr=0.1, d=3, n=100	0.6186	0.6853	0.5802	2.0399
Random Search	n=125	0.6372	0.6415	0.5838	3.0769
PSO	**n=96, lr=0.1718, d=3**	**0.6605**	**0.6883**	**0.6059**	**21.3423**

Source: (Research Results, 2025)

The effectiveness of the hyperparameter optimization stage is further evidenced through a comparative performance analysis of XGBoost under different tuning strategies. As illustrated in Figure 3, the PSO-optimized XGBoost model consistently outperforms Grid Search and Random Search across all evaluated metrics, including Accuracy, F1-Score, and ROC-AUC [25]. This performance uplift indicates that the PSO mechanism is able to more effectively explore the hyperparameter space and converge toward globally optimal solutions. In particular, the observed improvement in F1-Score highlights PSO's capability to enhance balanced classification performance under class-imbalanced conditions, thereby reinforcing its suitability as an optimization strategy for logistics SLA violation prediction [26].

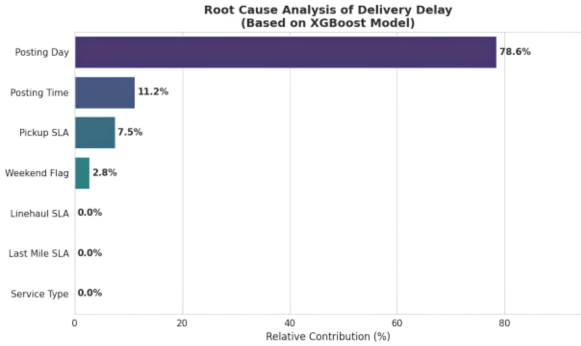


Source: (Research Results, 2025)

Figure 3. Performance Metrics Of Xgboost Across Different Tuning Methods

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8282

The fourth stage analyzes the interpretability of the best-performing model to move beyond predictive accuracy toward operational insight [27]. As shown in Figure 4, feature importance estimated using MDI criterion reveals a clear hierarchy of risk drivers for mail delivery delays. Posting Day is the primary determinant of mail delivery delay risk, contributing 78.6% of the total importance. This confirms that weekly temporal structure dominates delay behavior in mail delivery operations. Posting Time provides a secondary contribution of 11.2%, indicating that intra-day mail and dispatch timing further modulates delay risk with substantially lower impact. Pickup SLA contributes 7.5%, reflecting a limited but measurable effect of upstream mail collection commitments on delay occurrence. The Weekend Flag shows a minor influence (2.8%), suggesting that non-working days have a marginal role once posting-day effects are accounted for in mail delivery processes. In contrast, Linehaul SLA, Last Mile SLA, and Service Type exhibit negligible contributions (0.0%), indicating that these factors do not significantly affect mail delay prediction within the modeled operational context. This importance structure confirms that temporal dynamics outweigh service differentiation in driving delivery delay risk and demonstrates that the PSO-XGBoost model effectively captures the underlying operational mechanisms of logistics performance.



Source: (Research Results, 2025)

Figure 4. Feature importance ranking

The final stage of the research pipeline validates the proposed model through real-world risk scoring and prescriptive decision support. As shown in Table 3, the model assigns probabilistic risk scores that consistently distinguish Over SLA and On Time shipments, while maintaining correct validation outcomes. This result confirms that the PSO-optimized XGBoost model delivers interpretable and actionable predictions suitable for operational decision making.

Table 3 Operational Risk Synthesis: Predictive Scores and Prescriptive Action Plans

Receipt Number	Service	Post Hour	Risk (%)	Prediction	Action Plan	Validation
High Risk Sample						
P2507220002599	EC3	6	97.76	Over SLA	Priority Handling	Match
P2507190006140	EC3	23	79.27	Over SLA	Priority Handling	Match
P2509220181680	EC3	21	69.07	Over SLA	Priority Handling	Match
Low Risk Sample						
P2509250053199	PE	11	29.88	On Time	Standard Process	Match
P2507280056928	PE	11	33.50	On Time	Standard Process	Match
P2507230069643	PE	12	43.33	On Time	Standard Process	Match

Source: (Research Results, 2025)

Table 4 Comparation Result Learning

Author, citation	Type Dataset	Preprocessing			PSO	MDI	Model	Model Performance Result
		Missing Value Handling	Outlier Removal (IQR)	SMOTE				
Banjanin et al. [13]	Structured tabular data	✓	✗	✗	✗	✗	Random Forest	Accuracy = 0.95, F-Beta= 0,96, AUC = 0.99
Rezki & Mansouri [28]	Structured tabular data	✓	✗	✗	✗	✗	Gradient Boosting	Accuracy = 94%
Rokoss et al. [29]	Structured tabular data	✓	✗	✗	✗	✗	XGBoost	R^2 = 0.86
Sattar et al. [30]	Structured tabular data	✓	✗	✗	✗	✗	XGBoost	RMSE = 0.5333, MAPE = 0.48%, MAE = 0.1571, R^2 = 0.99999
Inaç et al. [31]	Data tabular	✗	✗	✗	✗	✗	Gradient Boosting	R^2 = 0.845
This Study	**Time-series**	✓	✓	✓	✓	✓	**XGBoost**	**Accuracy = 0.6605, ROC-AUC = 0.6883, F1-Score = 0.6059**

Source: (Research Results, 2025)

This study shows significant methodological differences compared to previous studies, especially in terms of data type and modeling approach. Most previous studies still rely on structured tabular data that represents static operational conditions whereas this study has adopted time series data based on operational temporal factors [13], [28], [29], [30], [31]. This approach allows the model to capture dynamic patterns such as daily and weekly cycles and calendar effects, thereby improving the level of data representation proximity to actual operational conditions. Thus, the prediction results obtained are more adaptive to time variations and logistics system complexity compared to the conventional approach used in previous studies.

In terms of model optimization, previous studies still show limitations in the application of hyperparameter tuning, which generally only uses Grid Search without exploring a wider parameter space [28]. Even in the delivery time determination case study, optimization did not include Random Search or metaheuristic approaches [13]. This study overcomes these limitations by comprehensively integrating Grid Search, Random Search, and PSO, thereby providing stronger empirical justification for the improvement in model performance. The evaluation results show that the best model in this study achieved a good performance value in the range of 66.04 although objectively there is still room for performance improvement in the future [13], [29].

In addition, the aspect of model interpretability in previous studies remains relatively limited, where feature importance analysis is reported only in a small number of works, such as a machine learning based investigation of delivery time estimation in small batch manufacturing environments and studies on supply chain optimization [13], [30]. This study expands on these contributions by systematically applying Mean Decrease Impurity (MDI) to identify the dominant factors causing SLA violations. However, this study still has limitations that can be further developed, including integrating additional explainability methods such as SHAP or Permutation Importance, exploring sequence learning-based models, and improving model generalization through cross-regional and cross-time period data expansion. With these developments, the performance and depth of predictive analysis are expected to be significantly improved in further research.

CONCLUSION

This study confirms that the integration of PSO with the XGBoost classifier provides an effective and interpretable framework for predicting logistics SLA violations. The proposed model captures temporal delay patterns from operational features and supports early identification of shipments with higher risk of exceeding SLA thresholds. Experimental results show that PSO-XGBoost achieves the most balanced predictive performance among the evaluated models, with an Accuracy of 0.6605, ROC-AUC of 0.6883, and F1-score of 0.6059. These results indicate that PSO-based hyperparameter optimization improves class-sensitive prediction performance under imbalanced logistics data. The study also shows that temporal features play a dominant role in mail delivery delay risk. Feature importance analysis identifies posting day as the strongest contributor, followed by posting time and pickup SLA, indicating that delay risk is mainly influenced by weekly and intra-day operational patterns. These findings provide practical insight for proactive SLA monitoring and operational prioritization. Future research should incorporate external contextual variables, such as weather conditions, traffic density, and regional operational characteristics, to improve predictive sensitivity. Further experiments may also apply SHAP or permutation-based importance measures, sequence-aware learning architectures, and multi-period cross-regional datasets to strengthen interpretability and model generalization.

REFERENCE

[1] C. Peoples, A. Moore, and N. Georgalas, “Service Level Agreement (SLA) Chains Supported by Cloud in a Complex Port Ecosystem with Competing Stakeholder Goals,” *EAI Endorsed Trans. Cloud Syst.*, p. 174394, 2022, doi: 10.4108/eai.26-7-2022.174394.

[2] O. A. Alamri, “Analysing service level agreements with multiple customers.,” RMIT University, 2024.

[3] R. E. R. Shawon, “Enhancing supply chain resilience across US regions using machine learning and logistics performance analytics,” *Int. J. Appl. Math.*, vol. 38, no. 4s, pp. 182–205, 2025.

[4] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, “Ensemble learning for disease prediction: A review,” in *Healthcare*, 2023, p. 1808, doi: 10.3390/healthcare11121808.

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8282

[5] W. Zhou and Z. Xie, "Enhancing Sealing Performance Predictions: A Comprehensive Study of XGBoost and Polynomial Regression Models with Advanced Optimization Techniques," *Materials (Basel).*, vol. 18, no. 10, p. 2392, 2025, doi: 10.3390/materials18102392.

[6] R. Garine and R. K. Chakrabortty, "A deep learning and policy optimization approach for supply chain order classification," *Supply Chain Anal.*, p. 100166, 2025, doi: 10.1016/j.sca.2025.100166.

[7] A. G. Gad, "Particle swarm optimization algorithm and its applications: A systematic review.," *Arch. Comput. methods Eng.*, vol. 29, no. 5, 2022, doi: 10.1007/s11831-021-09694.

[8] T. M. Shami, A. A. El-Saleh, M. Alswaitti, Q. Al-Tashi, M. A. Summakieh, and S. Mirjalili, "Particle swarm optimization: A comprehensive survey," *IEEE Access*, vol. 10, pp. 10031–10061, 2022, doi: 10.1109/ACCESS.2022.3142817.

[9] A. M. Musolf, E. R. Holzinger, J. D. Malley, and J. E. Bailey-Wilson, "What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics," *Hum. Genet.*, vol. 141, no. 9, pp. 1515–1528, 2022, doi: 10.1007/s00439-021-02402-z.

[10] H. Wang, Q. Liang, J. T. Hancock, and T. M. Khoshgoftaar, "Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods," *J. Big Data*, vol. 11, no. 1, p. 44, 2024, doi: 10.1186/s40537-024-00905-w.

[11] S. Fei, L. Li, Z. Han, Z. Chen, and Y. Xiao, "Combining novel feature selection strategy and hyperspectral vegetation indices to predict crop yield," *Plant Methods*, vol. 18, no. 1, p. 119, 2022, doi: 10.1186/s13007-022-00949-0.

[12] M. A. Khan, A. Kanwal, S. Abbas, F. Khan, and T. Whangbo, "Intelligent Model for Predicting the Quality of Services Violation.," *Comput. Mater. \& Contin.*, vol. 71, no. 2, 2022, doi: 10.32604/cmc.2022.023480.

[13] M. K. Banjanin, M. Stojčić, \DJor\dje Popović, D. An\djelković, G. Jauševac, and M. Husić, "Classification Machine Learning Models for Enhancing the Sustainability of Postal System Modules Within the Smart Transportation Concept," *Sustainability*, vol. 17, no. 19, p. 8718, 2025, doi: 10.3390/su17198718.

[14] A. Alabrah, "An improved CCF detector to handle the problem of class imbalance with outlier normalization using IQR method," *Sensors*, vol. 23, no. 9, p. 4406, 2023, doi: 10.3390/s23094406.

[15] S. Datta, C. Ghosh, and J. P. Choudhury, "Classification of imbalanced datasets utilizing the synthetic minority oversampling method in conjunction with several machine learning techniques," *Iran J. Comput. Sci.*, vol. 8, no. 1, pp. 51–68, 2025, doi: 10.1007/s42044-024-00207-7.

[16] M. M. H. Sizan *et al.*, "AI-Enhanced Stock Market Prediction: Evaluating Machine Learning Models for Financial Forecasting in the USA," *J. Bus. Manag. Stud.*, vol. 5, no. 4, pp. 152–166, 2023, doi: 10.46828/jbms.v5i4.8609.

[17] S. F. Pane, M. D. Sulistiyo, A. A. Gozali, and A. Adiwijaya, "PSO-Enhanced ensemble techniques for pandemic prediction and feature importance analysis," *Int. J. Adv. Intell. Informatics*, vol. 11, no. 4, pp. 653–666, 2025, doi: 10.26555/ijain.v11i4.2091.

[18] A.-H. M. Emara, G. Atteia, and J. H. Alkhateeb, "Fine Tuning Hyperparameters of Deep Learning Models Using Metaheuristic Accelerated Particle Swarm Optimization Algorithm," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.11087552.

[19] Y.-X. He, S.-H. Lyu, and Y. Jiang, "Interpreting deep forest through feature contribution and mdi feature importance," *ACM Trans. Knowl. Discov. Data*, 2024, doi: 10.1145/3641108.

[20] A. Agarwal, A. M. Kenney, Y. S. Tan, T. M. Tang, and B. Yu, "MDI+: A flexible random forest-based feature importance framework," *arXiv Prepr. arXiv2307.01932*, 2023, doi: 10.5281/zenodo.8111870.

[21] V. Da Poian *et al.*, "Exploratory data analysis (EDA) machine learning approaches for ocean world analog mass spectrometry," *Front. Astron. Sp. Sci.*, vol. 10, p. 1134141, 2023, doi: 10.3389/fspas.2023.1134141.

[22] N. Ekbote, P. Dhanshetti, and S. Sakhrekar, "TECHNIQUES OF EXPLORATORY DATA ANALYSIS," *Madhya Pradesh J. Soc. Sci.*, vol. 28, no. 2, 2023.

[23] V. Teodorescu and L. Obreja Brașoveanu, "Assessing the Validity of k-Fold Cross-Validation for Model Selection: Evidence from Bankruptcy Prediction Using Random Forest and XGBoost," *Computation*, vol. 13, no. 5, p. 127, 2025, doi: 10.3390/computation13050127.

[24] S. Ünalan, O. Günay, I. Akkurt, K. Gunoglu,

and H. O. Tekin, "A comparative study on breast cancer classification with stratified shuffle split and K-fold cross validation via ensembled machine learning," *J. Radiat. Res. Appl. Sci.*, vol. 17, no. 4, p. 101080, 2024, doi: 10.1016/j.jrras.2024.101080.

[25] J. C. Obi, "A comparative study of several classification metrics and their performances on data," *World J. Adv. Eng. Technol. Sci.*, vol. 8, no. 1, pp. 308–314, 2023, doi: 10.30574/wjaets.2023.8.1.0054.

[26] H. Shafa, "Integration Of Machine Learning and Advanced Computing For Optimizing Retail Customer Analytics," *Int. J. Bus. Econ. Insights*, vol. 2, no. 3, pp. 1–46, 2022, doi: 10.63125/p87sv224.

[27] A. K. Kalusivalingam, A. Sharma, N. Patel, and V. Singh, "Leveraging Random Forests and Gradient Boosting for Enhanced Predictive Analytics in Operational Efficiency," *Int. J. AI ML*, vol. 3, no. 9, 2022.

[28] N. Rezki and M. Mansouri, "Machine learning for proactive supply chain risk management: Predicting delays and enhancing operational efficiency," *Manag. Syst. Prod. Eng.*, vol. 32, no. 3, pp. 345–356, 2024, doi: 10.2478/mspe-2024-0033.

[29] A. Rokoss, M. Syberg, L. Tomidei, C. Hülsing, J. Deuse, and M. Schmidt, "Case study on delivery time determination using a machine learning approach in small batch production companies," *J. Intell. Manuf.*, vol. 35, no. 8, pp. 3937–3958, 2024, doi: 10.1007/s10845-023-02290-2.

[30] M. U. Sattar, V. Dattana, R. Hasan, S. Mahmood, H. W. Khan, and S. Hussain, "Enhancing Supply Chain Management: A Comparative Study of Machine Learning Techniques with Cost--Accuracy and ESG-Based Evaluation for Forecasting and Risk Mitigation," *Sustainability*, vol. 17, no. 13, p. 5772, 2025, doi: 10.3390/su17135772.

[31] H.Inaç, Y. E. Ayözen, A. Atalan, and C. Ç. Dönmez, "Estimation of postal service delivery time and energy cost with e-scooter by machine learning algorithms," *Appl. Sci.*, vol. 12, no. 23, p. 12266, 2022, doi: 10.3390/app122312266.

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri