Volume 11 Issue 1 Year 2026 | Page 358-369 ISSN: 2527-9866
*Received: **29-01-2026** | Revised: **15-02-2026** | Accepted: **26-02-2026***

A PSO-Optimized Stacking Ensemble with Hybrid SMOTE-NC and Tomek Links for Bid-Based Winning Prediction in Procurement Projects

Kokoh Handoko[1], Agus Purnomo[2], Erna Mulyati[3]
[1,2,3] University of Logistics and International Business (ULBI), Bandung, Indonesia,
e-mail: 211230056@std.ulbi.ac.id[1],aguspurnomo@ulbi.ac.id [2], ernamulyati@ulbi.ac.id[3]
*Correspondence: 211230056@std.ulbi.ac.id.

Abstract: *This research aims to establish a classification model for the prediction of procurement winning outcomes based on bid value and owner cost estimation data. The main challenge in procurement analysis lies in severe class imbalance and complex non-linear relationships among pricing and procurement attributes. The research object utilizes procurement tender data from PT Pos Indonesia, including project owner, owner cost estimation, bid value, and procurement method. The proposed approach integrates Hybrid SMOTE and Tomek Links for class balancing, regulation-driven feature engineering, and a stacking ensemble model optimized using Particle Swarm Optimization. The stacking framework combines Random Forest, Extra Trees, and Gradient Boosting as base learners. The experimental evaluation demonstrates that the proposed approach delivers the strongest performance, achieving an AUC of 0.92, an accuracy of 0.89, and an F1-Macro of 0.81, thereby surpassing all individual classifiers and homogeneous ensemble methods considered in this study. This study concludes that the hybrid optimization-based ensemble approach is effective for improving procurement winning prediction accuracy and provides a reliable decision-support tool for data-driven and regulation-compliant procurement processes.*

Keywords: *Procurement Winning Prediction, Stacking Ensemble Learning, Hybrid SMOTE–Tomek Links, Particle Swarm Optimization.*

I. INTRODUCTION

Procurement projects are one of the critical operations, both in public and private sectors, that determine the efficiency, transparency, and financial sustainability of large-scale expenditures [1]. Pricing strategies, along with the accuracy of cost estimation and compliance with regulatory constraints, have a significant bearing on the success or failure of a bid in competitive procurement environments. However, procurement winning outcome predictions remain an elusive task due to complex interactions among bid values, owner cost estimations, procurement methods, and organizational factors [2].

Classical rule based and descriptive procurement evaluation methods are limited in capturing nonlinear and hidden patterns in complex bid data, while machine learning based approaches have shown superior effectiveness in large-scale, high dimensional decision making domains [3]. As a result, data driven predictive models have been adopted to improve decision accuracy, with ensemble learning receiving particular attention for its ability to integrate multiple models and enhance robustness. Previous studies show that ensemble approaches such as stacking, Gradient Boosting, and Random Forest consistently outperform single classifiers by capturing complementary patterns missed by individual models [4]. Their performance remains sensitive to hyperparameter tuning and data imbalance.

This study focuses on price-related and procedural procurement attributes that are consistently available in historical tender data. While non-price factors such as technical compliance and vendor capability also influence outcomes, such information was unavailable in structured form; therefore, the proposed model is positioned as a decision-support tool for pricing and administrative evaluation rather than a comprehensive or universally generalizable procurement decision model. A combination of oversampling and undersampling methods, including the use of Tomek Links and the SMOTE algorithm, is effective in addressing the issues of the class boundaries and noisy data [5], [6], [7]. Apart from forecasting outcomes, modern machine learning tools also need to support optimization and interpretation. Hyperparameter optimization is an important ingredient in optimizing model performance, for which metaheuristic algorithms such as Particle Swarm Optimization (PSO) have shown significant efficacy in optimizing ensemble models for hyperparameters [8], [9]. In addition, interpretation of models through tools such as feature importance calculations via Mean Decrease in Impurity (MDI) can provide insights that can be essential in understanding the underlying factors that can affect predicted success in procurement activities [10].

Government prediction models require domain knowledge to ensure regulatory compliance. In Indonesia, Presidential Regulation No. 12 of 2021 mandates the automatic disqualification of bids exceeding the Owner's Cost Estimate (HPS), and this rule is incorporated into the data preprocessing pipeline solely to enforce bid eligibility and remove logically invalid outcomes, thereby improving data validity [11]. While such regulations define necessary conditions for participation, they do not fully determine winning outcomes, as many regulation-compliant bids still fail. For bids priced below the HPS, outcomes remain influenced by procedural, institutional, and competitive factors, motivating the use of machine learning to capture complex non-linear patterns beyond explicit rules. This study addresses limitations in prior procurement analytics by proposing a PSO-optimized stacking ensemble integrated with Hybrid SMOTE-NC and Tomek Links, regulation-aware preprocessing, and interpretable feature analysis for bid outcome prediction.

The key contributions of this study include the development of a hybrid ensemble learning framework that combines stacking-based classification with PSO for hyperparameter optimization and a Hybrid SMOTE-NC–Tomek Links strategy to address class imbalance in procurement datasets [12]. Domain-specific rule constraints derived from government procurement regulations are also integrated to ensure data validity and consistency. In addition, a comprehensive comparative evaluation of nine classification algorithms under various optimization schemes is conducted to identify the most effective predictive model, while Mean Decrease in Impurity analysis is applied to improve interpretability by highlighting the most influential factors affecting procurement outcomes.

II. METHODS

A research methodology workflow outlines the structured steps for developing, validating, and evaluating a predictive model. As shown in Figure 1, this study adopts an eight-stage framework for procurement tender winning classification, covering data collection, preprocessing with regulation-based validation, feature engineering with class-imbalance handling, model construction, hyperparameter tuning, feature importance analysis, performance evaluation, and a recommendation stage that translates predictions into regulation-compliant decision-support insights.

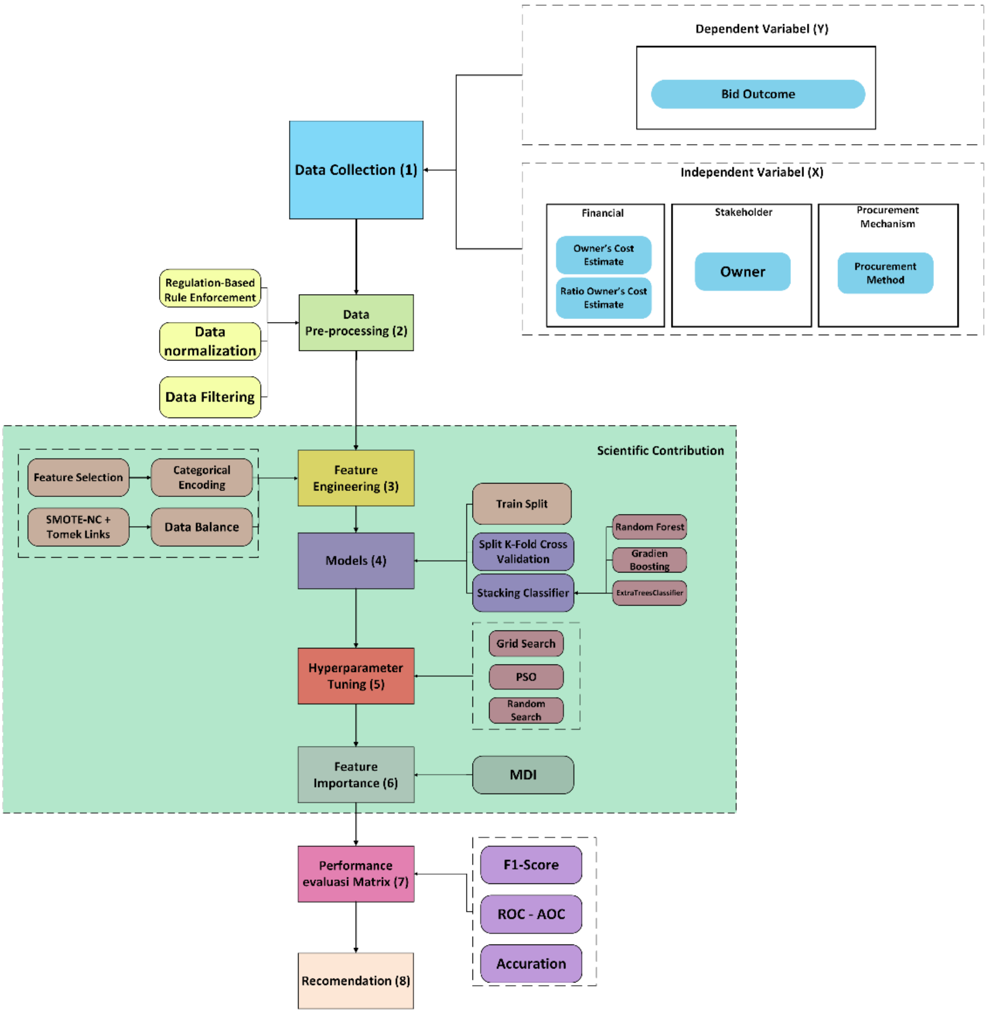

Figure 1. The proposed research methodology

A. Data Collection

This study uses a proprietary procurement tender dataset from PT Pos Indonesia consisting of 1,341 records stored in a structured spreadsheet capturing aspects of procurement decision-making, including financial attributes, stakeholder characteristics, and procurement mechanisms. Four independent variables are examined project owner, owner's cost estimation (HPS), bid value, and procurement method while the dependent variable is bid outcome (is_winning), indicating tender award status. Bid outcomes are derived from historical procurement decisions, with regulatory rules, specifically the automatic disqualification of bids exceeding the owner's estimate (HPS), applied during data validation to exclude invalid winning cases; bids priced below HPS do not guarantee success, indicating that the bid–HPS relationship constitutes a necessary but insufficient condition for determining outcomes. Prior Prior to modeling, the dataset is preprocessed, feature-engineered, and class-balanced to ensure consistency and reduce bias and overfitting risk [13]. with institutional permission from PT Pos Indonesia and full anonymization.

B. Data Pre-Processing

The data pre-processing stage ensures numerical consistency, regulatory compliance, and analytical validity prior to model development. It includes data type correction by converting monetary attributes from text to numerical format, normalization of selected variables such as transforming the pipeline indicator into a binary (1/0) form, and the application of regulation-driven constraints based on Indonesian procurement rules by labeling bids exceeding the owner's cost estimation (HPS) as losing outcomes (value 0) to enforce domain consistency rather than deterministic decision rules. As a result, the predictive task focuses on distinguishing winning and losing bids within regulation-compliant tenders, where outcomes are not governed by a single rule, while a derived feature, the HPS ratio (ratio_hps), is constructed to represent bidders' relative pricing strategies through the proportional relationship between bid value and the owner's estimate, formally defined as follow:

$$\text{ratio\_hps} = \frac{\text{bid value}}{\text{HPS}} \tag{1}$$

In Equation (1), ratio_hps is the bid to estimate ratio measuring bidding aggressiveness, with the bid value relative to the Owner's Cost Estimate (HPS). This formulation evaluates pricing behavior independent of absolute contract values while preserving relative competitiveness.

C. Feature Engineering

Feature engineering applies categorical encoding to convert non-numeric stakeholder attributes, particularly project owner, into numerical form, followed by feature selection of bid- and procurement-related attributes relevant to winning outcomes. Class imbalance is addressed using a hybrid SMOTE-NC and Tomek Links resampling strategy applied only to the training data after the train–test split and within each stratified k-fold cross-validation fold, while the hold-out test set remains unchanged to prevent information leakage. SMOTE-NC integrates a categorical distance measure based on the Value Difference Metric with continuous feature interpolation derived from the SMOTE algorithm.

$$D_{\text{cat}}(a,b) = \sum_{j=1}^{m} \left| \frac{n_{aj}^{(c)}}{n_a} - \frac{n_{bj}^{(c)}}{n_b} \right|^{\gamma} \tag{2}$$

In Equation (2), $\left(D_{\text{cat}}(a,b)\right)$ describes the categorical distance between two nominal feature values *(a)* and *(b)*. The term $\left(n_{aj}^{(c)}\right)$ denotes the number of occurrences of value *(a)* belonging to class *(j)*, while (n_a) represents the total occurrences of value *(a)* across all classes. Similarly, $\left(n_{bj}^{(c)}\right)$ and (n_b) correspond to value *(b)*. The parameter (γ) is a constant, typically set to 1, controlling the sensitivity of class-conditional differences [14]. Tomek Links are subsequently applied as a post-processing step to remove overlapping majority-class samples near class boundaries, thereby improving class separability and reducing noise.

D. Model

After preprocessing, feature construction, and hybrid imbalance handling, modeling targets robust generalization with minimal overfitting. The dataset is split 80:20 into development and hold-out sets for independent evaluation [15]. Stratified k-fold cross-validation with a fixed number of folds is applied to the training data to obtain stable performance estimates and avoid tuning bias. All resampling is confined to the training set, while evaluation is performed on an independent hold-out test set. A stacking classifier is selected for its ability to combine complementary base learners.
It integrates Random Forest, Extra Trees, and Gradient Boosting with a Logistic Regression meta-classifier, outperforming single and homogeneous ensembles [16], [17]. In addition to binary classification, the model produces a percentage-based risk score derived from the meta-classifier's predicted class probability, serving as a model-based confidence indicator of unfavorable outcome likelihood. Formally, the risk score is computed as:

$$Risk = \left(1 - P\left(is\_Winning\right)\right) \times 100 \tag{3}$$

In equation (3), Risk represents the relative risk score expressed as a percentage and the operator $\left(1-P\left(\text{is\_Winning}\right)\right)$ indicates the complement of the predicted winning probability. The multiplication by 100 scales the risk value into percentage form.

E. Hyperparameter Tuning

Hyperparameter tuning improves model performance and generalization by optimizing parameter configurations. This study applies Grid Search, Random Search, and Particle Swarm Optimization (PSO) to balance exhaustive evaluation, stochastic efficiency, and global search. PSO is a population-based method inspired by swarm behavior, where each particle represents a candidate configuration defined by its position $\left(z_i^{(k)}\right)$ and velocity $\left(u_i^{(k)}\right)$ at iteration *(k)*. he velocity and position update rules are formulated as follows:

$$u_i^{(k+1)} = \alpha u_i^{(k)} + \beta_1 \xi_1 \left(b_i^{(k)} - z_i^{(k)}\right) + \beta_2 \xi_2 \left(h^{(k)} - z_i^{(k)}\right) \quad (4)$$

$$z_i^{(k+1)} = z_i^{(k)} + u_i^{(k+1)} \quad (5)$$

Equation (4) defines the velocity update of particle (i) at iteration (k+1), where (\alpha) controls exploration–exploitation balance, (\beta_1) and (\beta_2) are cognitive and social acceleration factors, and (\xi_1, \xi_2 \in [0,1]) introduce stochasticity. The terms (b_i^{(k)}) and (h^{(k)}) denote the particle's personal best and the global best positions, respectively. Equation (5) describes the position update, where the new position (z_i^{(k+1)}) is obtained by adding the updated velocity (u_i^{(k+1)}) to the current position (z_i^{(k)}), enabling particles to move toward promising regions in the hyperparameter space [18], [19].

F. Feature Important

Feature importance analysis improves interpretability and identifies influential variables, with Mean Decrease in Impurity (MDI) applied to tree-based models by aggregating impurity reductions across all splits. The MDI for feature (j) is computed as follows.:

$$\text{MDI}(j) = \frac{1}{T}\sum_{s=1}^{T} \Delta I(s)\, , 1_{s \text{ ue } j} \quad (6)$$

In Equation (6), T denotes the total number of trees in the ensemble, s represents a split node, and $\left(\Delta \text{I}(s)\right)$ indicates the impurity reduction at split (s), measured using Gini impurity or entropy depending on the model. The indicator function $\left(1_{s \in j}\right)$ equals 1 when feature j is used at split s and 0 otherwise [20]. This formulation averages impurity reductions across all trees, providing a global measure of feature contribution to the ensemble's predictive accuracy.

G. Evaluation Metrics

Model evaluation uses multiple classification metrics to assess predictive performance and class separation under imbalanced procurement data. This study employs Accuracy, F1-Macro, and ROC–AUC as complementary measures to capture overall correctness and class discrimination. Accuracy is defined as the proportion of correctly predicted instances and is expressed as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

In Equation (7), (TP) denotes true positives, (TN) true negatives, (FP) false positives, and (FN) false negatives [21]. Accuracy reflects overall correctness but is inadequate for imbalanced classes, so F1-Macro is used as the harmonic means of precision and recall to balance false positive and false negative errors and is computed as follows:

$$\text{F1-Macro} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

In Equation (8) Precision and recall are defined as $\left(Precision = \frac{TP}{TP+FP}\right)$ and $\left(Recall = \frac{TP}{TP+FN}\right)$, respectively [21]. The F1-Macro score is used as the primary F1-based metric, computed as the unweighted mean of class-wise F1 scores to ensure equal importance for winning and non-winning outcomes under class imbalance. Model discrimination is assessed using ROC–AUC, which measures class separability across decision thresholds, with the ROC curve plotting the true positive rate (TPR) against the false positive rate (FPR), defined as follows:

$$\text{TPR} = \frac{TP}{TP+FN}, \text{FPR} = \frac{FP}{FP+TN} \quad (9)$$

In Equation (9) The Area Under the Curve (AUC) is computed as the integral of the ROC curve and represents the probability that the model assigns a higher score to a randomly chosen positive instance than to a randomly chosen negative instance [22]. A higher ROC–AUC value indicates stronger discriminative performance independent of classification thresholds.

III. RESULTS AND DISCUSSION

A. Result

This section presents the experimental results of the PSO-optimized stacking ensemble. From 1,340 procurement records, preprocessing removed 87 invalid entries, yielding 1,253 regulation-compliant instances for model development and evaluation. [23]. Within this final dataset, the target variable is_winning exhibited a dominant positive class, with 1,073 winning instances (1) and 180 non-winning instances (0), indicating substantial class imbalance and justifying the application of imbalance handling techniques in subsequent modeling stages [24].

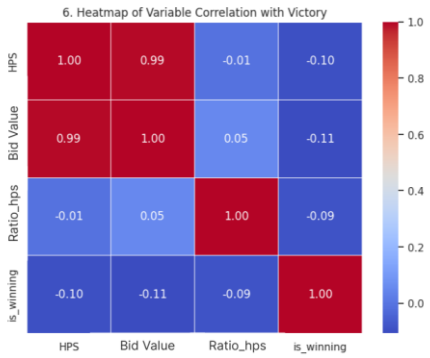

Figure 2. *Heatmap Correlation*

Figure 2 shows a very strong positive correlation between HPS and bid value (0.99), while is_winning has weak negative correlations with HPS (−0.10), bid value (−0.11), and ratio_hps (−0.09), and ratio_hps is weakly correlated with HPS (−0.01) and bid value (0.05). These findings indicate that procurement outcomes are not linearly driven by pricing variables alone, supporting non-linear ensemble learning. Data preparation improves interpretability through normalization and SMOTE to address class imbalance [25] [26]. The extremely high correlation between HPS and bid value reflects inherent pricing dependency handled by tree-based ensemble learners using recursive partitioning. Accordingly, HPS and bid value act as joint influences on decision boundaries, while ratio_hps captures pricing competitiveness and reduces redundant linear influence.

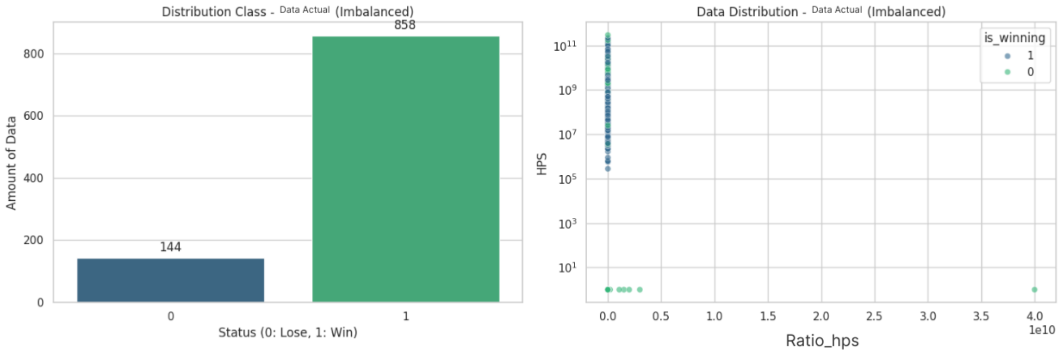

Figure 3. Class distribution and feature space of the original imbalanced dataset.

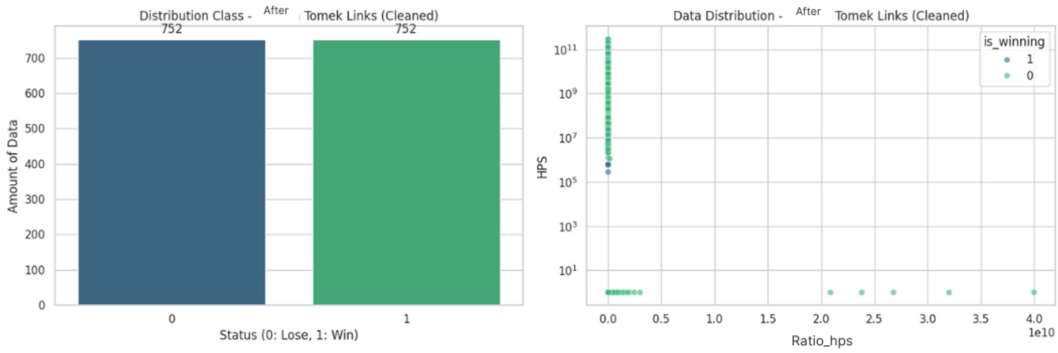

Figure 4. Balanced class distribution and feature space after Hybrid SMOTE–Tomek Links resampling.

Figures 3 and 4 show the effect of the Hybrid SMOTE–Tomek Links method on class distribution. Prior to resampling, the dataset was highly imbalanced, with 858 winning instances (1) and 144 non-winning instances (0), while after resampling it became fully balanced with 752 instances in each class. The post-resampling plots display a more uniform distribution across the feature space, indicating effective removal of overlapping and borderline samples and directly supporting the objective of reducing majority-class bias in procurement winning prediction. This study shows that stratified k-fold cross-validation yields varying predictive performance across models, with the Stacking Classifier consistently outperforming individual base classifiers in F1-Macro score, as shown in Figure 5. Performance remains stable for k values from 5 to 10, indicating robust generalization independent of performance-driven k selection. All F1-Macro scores are obtained from stratified k-fold cross-validation on training data after hybrid SMOTE–Tomek Links resampling.

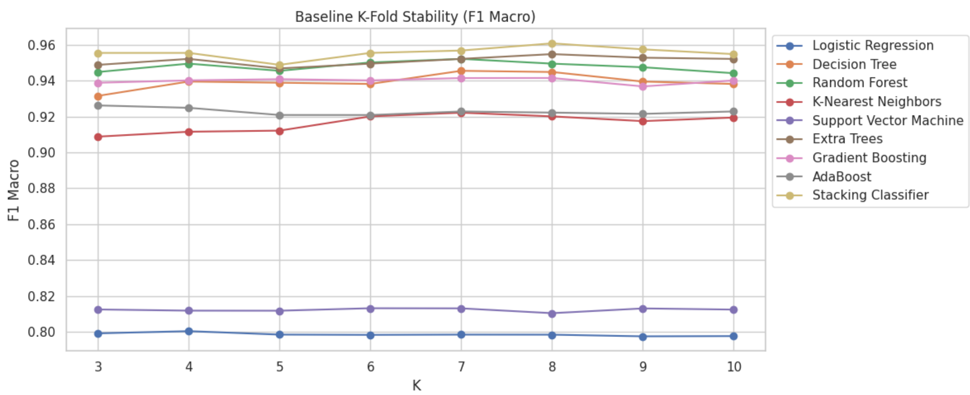

Figure 5. K-Fold Cross Validation model

Table 1. Result Model after Hyperparameter Tuning

Nama Model	Metode Tuning	F1-Macro	Accuracy	ROC-AUC
Logistic Regression	Grid Search	0.633320302	0.717131474	0.701033592
Decision Tree	Grid Search	0.78831243	0.876494024	0.861369509
Random Forest	Grid Search	0.784489683	0.876494024	0.915891473
K-Nearest Neighbors	Grid Search	0.698219994	0.836653386	0.761046512
Support Vector Machine	Grid Search	0.625149343	0.721115538	0.738501292
Extra Trees	Grid Search	0.785543404	0.880478088	0.910465116
Gradient Boosting	Grid Search	0.784489683	0.876494024	0.912855297
AdaBoost	Grid Search	0.741617647	0.832669323	0.898966408
Stacking Classifier	Grid Search	0.80880237	0.892430279	0.919767442
Logistic Regression	PSO	0.633320302	0.717131474	0.700904393
Decision Tree	PSO	0.765113232	0.860557769	0.879521964
Random Forest	PSO	0.78831243	0.876494024	0.917312661
K-Nearest Neighbors	PSO	0.698219994	0.836653386	0.761046512
Support Vector Machine	PSO	0.696419932	0.84063745	0.785529716
Extra Trees	PSO	0.770585791	0.868525896	0.907299742
Gradient Boosting	PSO	0.803555456	0.888446215	0.916020672
AdaBoost	PSO	0.73804251	0.836653386	0.904134367
Stacking Classifier	**PSO**	**0.80880237**	**0.892430279**	**0.92002584**
Logistic Regression	Random Search	0.633320302	0.717131474	0.701033592
Decision Tree	Random Search	0.765113232	0.860557769	0.879521964
Random Forest	Random Search	0.779534475	0.87250996	0.917635659
K-Nearest Neighbors	Random Search	0.65935466	0.804780876	0.755813953
Support Vector Machine	Random Search	0.625149343	0.721115538	0.740568475
Extra Trees	Random Search	0.760997905	0.860557769	0.90497416
Gradient Boosting	Random Search	0.789523703	0.880478088	0.903875969
AdaBoost	Random Search	0.73804251	0.836653386	0.903875969
Stacking Classifier	Random Search	0.80880237	0.892430279	0.919767442

The results show that hyperparameter optimization strongly affects model performance across classifiers. Three strategies PSO, Grid Search, and Random Search are evaluated, with PSO consistently achieving the best results in Table 2, including a ROC-AUC of 0.92, accuracy of 0.89, and F1-Macro of 0.80. These findings highlight the effectiveness of PSO in identifying optimal hyperparameter configurations and support the novelty of integrating PSO-based optimization within a stacking ensemble framework for procurement outcome prediction.

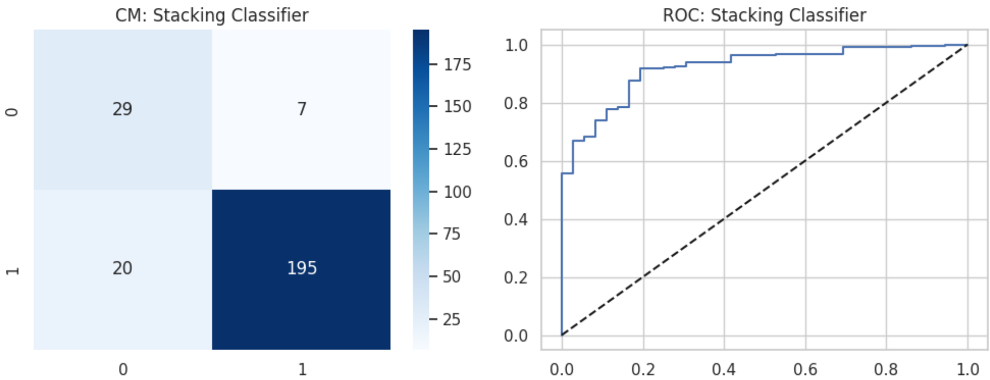

Figure 6. Confusion matrix and ROC curve of the stacking classifier.

Figure 6 shows the robustness of the Stacking Classifier, with only 20 false negatives, 7 false positives, and an ROC–AUC above 0.90, indicating strong class separability and reliable decision-support capability with minimal regulatory and financial risk.

Table 2. Predicted Tender Win Rates, Risk Levels, and Model Validation across Various Procurement Segments

Segment	Procurement Method	Owner's Estimate	Offer	Prediction	Risk (%)	Validation
A. Predicted *is_win* Sample						
Government	Tender	605285270	508182673.1	is_win	4	match
BUMN	Tender	24300000000	24300000000	is_win	43	match
Government	Tender	80000000	78000000	is_win	18	Match
B. Predicted *is_lose* Sample						
Private Company	Tender	200000000	200000000	is_lose	47.2	match
Private Company	Tender	3000000000	3000000000	is_lose	7.5	match
Government	Tender	6428124000	6428124000	is_lose	11.4	match

Table 2 shows consistent prediction of winning and losing outcomes. Segment A is predicted as *is_win* with low–moderate risk (4%–43%), while segment B is predicted as *is_lose* with higher risk variation (7.5%–47.2%), confirming strong discriminative capability based on the generated risk scores.

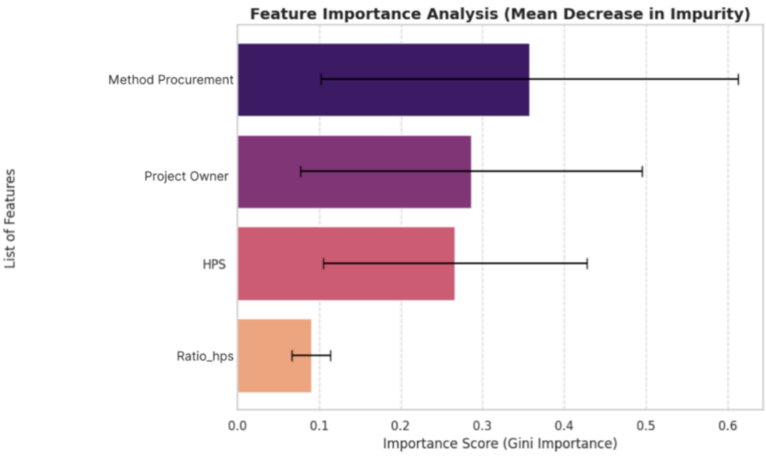

Figure 7. Feature importance ordering derived from the mean decrease in impurity (MDI) metric.

Feature importance is evaluated using Mean Decrease in Impurity (MDI). As shown in Figure 7, Metode Pengadaan is most influential (0.36), followed by Project Owner (0.29) and HPS (0.27), while rasio_hps has minimal impact (0.08). These results highlight the dominant role of procurement methods and project ownership in shaping classification boundaries. Given the strong HPS bid correlation, important scores reflect structural relevance rather than causality. Feature importance supports qualitative interpretation, with procurement method and project ownership aligning with Indonesian procurement practices.

B. Discussion

The proposed hybrid framework improves predictive performance but remains a decision-support tool operating within regulation-compliant procurement. As the analysis relies on data from a single institution (PT Pos Indonesia) without external or temporal validation, generalization beyond the observed scope is limited. The PSO-optimized stacking ensemble improves accuracy, handles class imbalance, and offers interpretability, while computational complexity, reliance on numerical pricing features, and base-learner dependency may limit scalability and robustness.

Future research may extend this work by incorporating qualitative bid evaluation criteria, multi-institutional datasets, and real-time procurement data to enhance scalability and practical relevance. As summarized in Table 2, this study contributes to the tender-winning prediction literature by jointly integrating hybrid class imbalance handling using SMOTE–Tomek Links, stacking-based ensemble learning, regulation-aware preprocessing, and feature importance analysis within a single procurement-focused framework. Prior studies generally apply these components in isolation, which limits their effectiveness under complex and imbalanced procurement conditions. In contrast, this study using PT Pos Indonesia data uniquely combines hybrid resampling, stacking classifiers, and feature importance evaluation in one framework. Existing works either apply SMOTE without Tomek Links, as in Figueroa-Gómez and Galpi [27], or do not employ hybrid resampling to address overlapping and borderline instances [28], [29], [30], [31] while most rely on single classifiers without stacking, with the exception of Cernăzanu-Glăvan and Bulzan [28], who did not incorporate hybrid imbalance handling. As a result, the proposed approach achieves balanced performance with accuracy, F1 score, and ROC-AUC of approximately 0.89, 0.81, and 0.92, respectively, and further distinguishes itself by incorporating feature importance analysis, which is largely absent in previous studies.

IV. CONCLUSIONS

This study shows that the PSO-optimized stacking ensemble effectively classifies procurement winning outcomes within the evaluated institutional context. The stacking model combining Random Forest, Extra Trees, and Gradient Boosting achieves an ROC–AUC of 0.92, accuracy of 0.89, and F1 score of 0.81, outperforming single and homogeneous models and alternative tuning strategies. These results confirm the effectiveness of heterogeneous ensemble stacking with PSO optimization and Hybrid SMOTE–Tomek Links in handling class imbalance and capturing non-linear bid relationships. Future work may improve generalizability and deployment readiness by incorporating multi-institutional and temporal datasets, qualitative and non-price attributes, and scalability enhancements through lightweight ensembles, online learning, and advanced interpretability methods such as permutation importance or SHAP.

REFERENCES

[1] M. A. Berawi, M. Sari, N. H. A. Amiri, suci indah Susilowati, S. R. Utami, and A. Kulachinskaya, "Developing a Machine Learning Model to Improve the Accuracy of Owner Estimate Cost in the Capital Expenditure Procurement Process," *Int. J. Technol.*, vol. 16, no. 4, p. 1179, Jul. 2025, doi: 10.14716/ijtech.v16i4.7409.

[2] P. Marvan and V. Hromádka, "From Price to Value: Implementing Best Value Procurement in the Czech Public Sector—A Case Study with Survey Insights," *Buildings*, vol. 15, no. 21, p. 3981, Nov. 2025, doi: 10.3390/buildings15213981.

[3] B. L. Oo, A. T. Nguyen, Y. Ahn, and B. T. H. Lim, "Predicting the number of bidders in construction competitive bidding using explainable machine learning models," *Constr. Innov.*, vol. 25, no. 7, pp. 158–188, Dec. 2025, doi: 10.1108/CI-10-2024-0325.

[4] O. Mermer, E. Zhang, and I. Demir, "A Comparative Study of Ensemble Machine Learning and Explainable AI for Predicting Harmful Algal Blooms," *Big Data Cogn. Comput.*, vol. 9, no. 5, p. 138, May 2025, doi: 10.3390/bdcc9050138.

[5] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.

[6] P. Branco, L. Torgo, and R. P. Ribeiro, "A Survey of Predictive Modeling on Imbalanced Domains," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1–50, Jun. 2017, doi: 10.1145/2907070.

[7] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, Oct. 2018, doi: 10.1016/j.neunet.2018.07.011.

[8] 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*, vol. 13, pp. 134506–134518, 2025, doi: 10.1109/ACCESS.2025.3591403.

[9] D. Zheng, C. Qin, and P. Liu, "Adaptive Particle Swarm Optimization Algorithm Ensemble Model Applied to Classification of Unbalanced Data," *Sci. Program.*, vol. 2021, pp. 1–13, Oct. 2021, doi: 10.1155/2021/7589756.

[10] C. Molnar, G. Casalicchio, and B. Bischl, "Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges," 2020, pp. 417–431. doi: 10.1007/978-3-030-65965-3_28.
[11] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and explainability of artificial intelligence in medicine," *WIREs Data Min. Knowl. Discov.*, vol. 9, no. 4, Jul. 2019, doi: 10.1002/widm.1312.
[12] Y. Sun, G. Li, N. Zhang, Q. Chang, J. Xu, and J. Zhang, "Development of ensemble learning models to evaluate the strength of coal-grout materials," *Int. J. Min. Sci. Technol.*, vol. 31, no. 2, pp. 153–162, Mar. 2021, doi: 10.1016/j.ijmst.2020.09.002.
[13] J. Zhang, X. Chu, A. Simeone, and P. Gu, "Machine learning-based design features decision support tool via customers purchasing data analysis," *Concurr. Eng.*, vol. 29, no. 2, pp. 124–141, Jun. 2021, doi: 10.1177/1063293X20963313.
[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
[15] H. Bichri, A. Chergui, and M. Hain, "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 2, 2024, doi: 10.14569/IJACSA.2024.0150235.
[16] D. E. Pangestika, A. Fitrianto, and K. Sadik, "Classification Performance of Stacking Ensemble with Meta-Model of Categorical Principal Component Logistic Regression on Food Insecurity Data," *Sci. J. Informatics*, vol. 11, no. 4, pp. 1085–1098, Mar. 2025, doi: 10.15294/sji.v11i4.15315.
[17] S. Q. Sultan, N. Javaid, N. Alrajeh, and M. Aslam, "Machine Learning-Based Stacking Ensemble Model for Prediction of Heart Disease with Explainable AI and K-Fold Cross-Validation: A Symmetric Approach," *Symmetry (Basel).*, vol. 17, no. 2, p. 185, Jan. 2025, doi: 10.3390/sym17020185.
[18] 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, p. 653, Nov. 2025, doi: 10.26555/ijain.v11i4.2091.
[19] K. Tang and C. Meng, "Particle Swarm Optimization Algorithm Using Velocity Pausing and Adaptive Strategy," *Symmetry (Basel).*, vol. 16, no. 6, p. 661, May 2024, doi: 10.3390/sym16060661.
[20] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," *Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, [Online]. Available: https://papers.nips.cc/paper_files/paper/2013/hash/e3796ae838835da0b6f6ea37bcf8bcb7-Abstract.html
[21] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, p. 6086, Mar. 2024, doi: 10.1038/s41598-024-56706-x.
[22] M. Imani and H. R. Arabnia, "Hyperparameter Optimization and Combined Data Sampling Techniques in Machine Learning for Customer Churn Prediction: A Comparative Analysis," *Technologies*, vol. 11, no. 6, p. 167, Nov. 2023, doi: 10.3390/technologies11060167.
[23] A. K. Haghighat, "Transportation Safety Enhancement by Implementing Computer Vision Based Models in Intelligent Transportation Systems," 2022, *Iowa State University*. [Online]. Available: https://dr.lib.iastate.edu/server/api/core/bitstreams/72f89f1e-6fa6-4a48-afb4-a0c9f7625a61/content
[24] V. Werner de Vargas, J. A. Schneider Aranda, R. dos Santos Costa, P. R. da Silva Pereira, and J. L. Victória Barbosa, "Imbalanced data preprocessing techniques for machine learning: a systematic mapping study," *Knowl. Inf. Syst.*, vol. 65, no. 1, pp. 31–57, Jan. 2023, doi: 10.1007/s10115-022-01772-8.
[25] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, "Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks," *Appl. Sci.*, vol. 13, no. 6, p. 4006, Mar. 2023, doi: 10.3390/app13064006.
[26] A. J. Mohammed, M. Muhammed Hassan, and D. Hussein Kadir, "Improving classification performance for a novel imbalanced medical dataset using SMOTE method," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3161–3172, 2020, [Online]. Available: https://eprints.cihanuniversity.edu.iq/id/eprint/971/
[27] Y. Figueroa-Gómez and I. Galpin, "Identifying Public Tenders of Interest Using Classification Models: A Comparative Analysis," *SN Comput. Sci.*, vol. 5, no. 1, p. 87, Dec. 2023, doi:

10.1007/s42979-023-02443-3.

[28] C. Cernăzanu-Glăvan and A.-Ștefan Bulzan, "Explainable AI and Fuzzy Linguistic Interpretation for Enhanced Transparency in Public Procurement: Analyzing EU Tender Awards," *Mathematics*, vol. 13, no. 13, p. 2215, Jul. 2025, doi: 10.3390/math13132215.

[29] R. Nai, R. Meo, G. Morina, and P. Pasteris, "Public tenders, complaints, machine learning and recommender systems: a case study in public administration," *Comput. Law Secur. Rev.*, vol. 51, p. 105887, Nov. 2023, doi: 10.1016/j.clsr.2023.105887.

[30] M. Y. Putra, R. N. Hayat, A. Chusyairi, D. I. Putri, and Solikin, "Machine Learning Model Based on REST API for Predicting Tenders Winner," in *2022 Seventh International Conference on Informatics and Computing (ICIC)*, IEEE, Dec. 2022, pp. 01–05. doi: 10.1109/ICIC56845.2022.10006976.

[31] Y. Torres-Berru and V. F. López Batista, "Data Mining to Identify Anomalies in Public Procurement Rating Parameters," *Electronics*, vol. 10, no. 22, p. 2873, Nov. 2021, doi: 10.3390/electronics10222873.