Jurnal Kecerdasan Buatan dan Teknologi Informasi
Vol. 5, No. 3, September 2026, pp. 809–820
e-ISSN: 2964-2922, p-ISSN: 2963-6191

Open access

Original Research

Employee Performance Classification Using Optimized Random Forest with Max Depth and N-Estimators at PT Bringin Gigantara

M. Imam Saputra [a], Rendra Gustriansyah [b*], Dwi Asa Verano [c]

[a,b,c]Department of Informatics, Faculty of Computer Science and Science, Universitas Indo Global Mandiri, Palembang, Indonesia

*Corresponding Author: rendra@uigm.ac.id

ABSTRACT

Employee performance assessment is a crucial component of human resource management because it supports strategic decisions such as contract renewal, incentives, and competency development. This study proposes an optimized Random Forest model for employee performance classification at PT Bringin Gigantara by tuning the `max_depth` and `n_estimators` hyperparameters. The dataset consists of monthly performance evaluation records from 100 contract employees during January–December 2024, including ten performance variables. During model development, hyperparameter optimization was conducted exclusively on the training set using Stratified 5-Fold Cross-Validation to evaluate parameter combinations and select the most stable and accurate configuration. The optimization results indicate that the best-performing configuration was `n_estimators` = 50 and `max_depth` = 10, achieving a mean cross-validation accuracy of 98%. The optimized model achieved 98.75% accuracy on the training set and 95% accuracy on the test set, with corresponding precision, recall, and F1-score values demonstrating good classification performance. The findings indicate that the optimized Random Forest model can provide a practical, objective, and data-driven decision-support tool for human resource personnel in evaluating employee performance and supporting contract renewal decisions. Nevertheless, the model is intended to support, rather than replace, managerial judgment.

How to Cite:
M. I. Saputra, R. Gustriansyah, and D. A. Verano, "Employee Performance Classification Using Optimized Random Forest with Max Depth and N-Estimators at PT Bringin Gigantara," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 5, no. 3, pp. 809–820, 2026.
Doi: `https://doi.org/10.69916/jkbti.v5i3.602`

KEYWORDS

Random Forest
Employee Performance
Machine Learning
Classification
N-Estimators

ARTICLE INFO

Received: July 24, 2026
Accepted: August 28, 2026
Available Online: September 23, 2026

1 INTRODUCTION

Employee performance appraisal is a critical component of human resource management because it supports strategic decisions such as contract renewal, incentives, promotion, and competency development. Reliable performance data help organizations maintain productivity and support long-term business sustainability, making objective and accurate evaluation essential [1].

In practice, however, manual and subjective performance assessments can result in bias, inconsistency, and limited transparency. The growing volume of evaluation data across multiple indicators further complicates conventional analysis, leaving potentially valuable performance patterns underutilized [2].

The advancement of data mining and machine learning provides a more systematic and objective alternative to conventional evaluation methods [3]. As a branch of artificial intelligence, machine learning can identify patterns in historical data and build predictive classification models, helping organizations uncover relationships among evaluation variables and support more consistent, data-driven decisions [4].

Random Forest is one of the most widely used machine learning algorithms for classification tasks. As an ensemble method, it builds multiple decision trees and aggregates their predictions, improving accuracy and

reducing the risk of overfitting associated with individual decision trees [5]. It also handles high-dimensional data effectively and generally provides good generalization to unseen data, making it suitable for employee performance classification involving multiple evaluation indicators [6].

Nonetheless, the performance of Random Forest is strongly influenced by its hyperparameter configuration. The `max_depth` parameter affects model complexity and its ability to capture data patterns, whereas `n_estimators`, which determines the number of trees, influences prediction stability and accuracy [7]. Inappropriate parameter settings may lead to underfitting or overfitting, highlighting the need for hyperparameter optimization to identify an effective combination [8].

Recent studies have applied machine learning to employee performance classification, but they differ in modeling strategy and optimization focus. Patel et al. compared Random Forest (RF), artificial neural network, decision tree, and XGBoost classifiers and proposed an ensemble model for employee-performance classification [12]. Their contribution focused primarily on combining several classifiers rather than systematically optimizing the internal RF hyperparameters. Tanasescu et al. investigated multiple machine learning algorithms for employee-performance prediction and incorporated hyperparameter tuning, demonstrating the value of model optimization in human resource analytics [4]. However, their study considered a broader multi-algorithm optimization framework rather than a focused analysis of the joint effects of `max_depth` and `n_estimators` in RF for employee contract-renewal classification [5]. In a different classification domain, Thomas and Kaliraj [6] demonstrated that systematic RF hyperparameter tuning can improve predictive performance, while Alonso-Sarría et al. showed that exhaustive hyperparameter optimization and cross-validation can reveal both accuracy and stability differences among RF configurations. These studies indicate that hyperparameter optimization is important, but they do not specifically address the joint optimization of `max_depth` and `n_estimators` for employee performance classification linked to contract renewal decisions [13]. Therefore, the novelty of this study lies in systematically evaluating these two RF hyperparameters using Stratified 5-Fold Cross-Validation and selecting a configuration that balances validation performance and generalization for the specific employee performance classification problem at PT Bringin Gigantara [14].

PT Bringin Gigantara, a company engaged in workforce provision and management, maintains employee performance evaluation data covering various assessment indicators [15]. These data provide an opportunity for data-driven analysis and decision support in human resource management. Recent international research has likewise shown that machine learning and data analytics can be used to make employee appraisal more systematic and less dependent on subjective judgment [16]. However, the specific use of an optimized Random Forest model to classify contract renewal status from employee performance indicators remains insufficiently explored in the context of the present case study [17].

Based on this background and the identified research gap, the present study applies the Random Forest algorithm to employee performance evaluation data from PT Bringin Gigantara [18]. The study systematically evaluates multiple combinations of `max_depth` and `n_estimators` using Stratified 5-Fold Cross-Validation to identify the optimal configuration. The resulting model classifies employee performance into the "Renewed" and "Not Renewed" categories [19]. The contribution of this study is therefore twofold: methodologically, it provides a focused and reproducible optimization of two influential Random Forest hyperparameters; practically, it provides a data-driven decision-support model that can assist human resource personnel in evaluating employee performance and supporting contract renewal decisions [20].

2 RESEARCH METHODS

2.1 Research Methodology

This study employs a quantitative approach to construct a predictive classification model for employee contract renewal decisions. The research workflow begins with the collection of employee performance evaluation data, followed by preprocessing that includes data cleaning, normalization, categorical encoding, and feature selection to ensure data quality and compatibility. The refined dataset is then divided into training and test sets to build the model and evaluate its performance on unseen data. Hyperparameter optimization is subsequently applied to the Random Forest algorithm by systematically tuning `n_estimators` and `max_depth`. Stratified 5-Fold Cross-Validation is used during hyperparameter optimization on the training set, while the test set is reserved for final evaluation. Finally, the optimized model is evaluated using a confusion matrix and standard metrics, including accuracy, precision, recall, and F1-score, as shown in Figure 1.

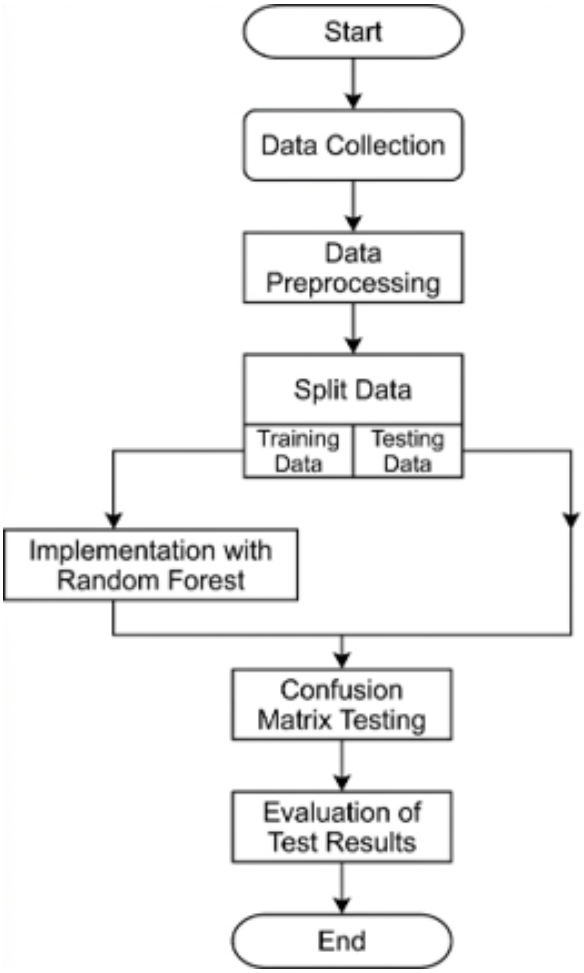

Figure 1. Research Process Flowchart

a Software Environment

The model was implemented in Python 3.11 using scikit-learn 1.5. Data preprocessing and analysis were performed using Pandas and NumPy, with Google Colab serving as the experimental environment. Random Forest and StratifiedKFold were implemented through scikit-learn. The use of a fixed `random_state` = 42 and `shuffle`=True in cross-validation provides a reproducible data partitioning procedure. This standardized computational environment and configuration support reproducibility of the proposed machine learning framework.

2.2 Data Collection and Variable Selection

This study uses a secondary dataset comprising historical employee performance evaluations provided by PT Bringin Gigantara. The dataset contains performance indicators used by the company to determine employee contract renewal decisions. The raw data were obtained from corporate records and structured for model training, hyperparameter optimization, and classification using the Random Forest algorithm.

2.3 Data Preprocessing

Data preprocessing is a critical phase for ensuring data quality and preparing the dataset for Random Forest classification, particularly when evaluating variations in `n_estimators` and `max_depth`. The raw employee evaluation data extracted from the company's internal system contained noise and inconsistencies that could introduce bias and reduce model accuracy. Therefore, a systematic data preparation pipeline was implemented, including missing-value handling, categorical encoding, feature scaling, and data cleaning, to transform the raw data into a clean, structured, and standardized format suitable for machine learning analysis.

a Selection of Random Forest Algorithm

Random Forest was selected as the primary classification algorithm because the employee performance dataset contains multiple assessment variables and a binary target, while the method can model nonlinear relationships and interactions among predictors without requiring a single decision-tree structure. Its ensemble mechanism combines predictions from multiple decision trees, which can improve predictive stability and reduce the variance associated with an individual tree. Random Forest was therefore considered appropriate for this study because

it can handle multiple performance indicators and provides a practical baseline for systematic hyperparameter optimization. The study does not claim that Random Forest is universally superior to all other classification algorithms; rather, it evaluates whether optimizing `n_estimators` and `max_depth` can produce a reliable model for the specific employee classification task.

2.4 Application of Random Forest

a Cross-Validation and Parameter Tuning

A Stratified 5-Fold Cross-Validation approach was employed to systematically select the optimal combination of `n_estimators` (number of decision trees) and `max_depth` (maximum tree depth) for the Random Forest model. The procedure was implemented using StratifiedKFold with `n_splits` = 5, `shuffle`=True, and `random_state` = 42. For each parameter combination, the training set was divided into five stratified folds; four folds were used for training and the remaining fold for validation, and this process was repeated until every fold had served as the validation fold. Stratification preserves the proportion of the "Renewed" and "Not Renewed" classes in each fold. Importantly, the independent test set was not used during hyperparameter optimization, thereby preventing data leakage. The mean validation accuracy across the five folds was used to compare candidate parameter combinations and select the optimal configuration.

In each iteration, four folds are used for training, while the remaining fold serves as the validation set. For each fold i, classification accuracy is computed as follows:

$$\text{Accuracy}_i = \frac{\text{TP}_i + \text{TN}_i}{\text{TP}_i + \text{TN}_i + \text{FP}_i + \text{FN}_i} \tag{1}$$

Here, TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. To determine the most stable parameter configuration, the mean cross-validation accuracy (CV_{mean}) is calculated across the five validation folds:

$$\text{CV}_{\text{mean}} = \frac{1}{5}\sum_{i=1}^{5} \text{Accuracy}_i \tag{2}$$

Evaluating different `n_estimators` values helps balance predictive stability and computational cost. A larger number of trees generally provides a more stable ensemble prediction, but it also increases computational requirements. The hyperparameter configuration with the highest CV_{mean} is selected as the optimal model configuration.

Stratified 5-Fold Cross-Validation was selected because the dataset contains an imbalance between the "Renewed" and "Not Renewed" classes. Stratification ensures that each fold preserves the class distribution of the original training set, producing more reliable and less biased validation results. Using five folds also allows the model to be trained and validated across multiple partitions while retaining sufficient data for training in each iteration.

b Parameter Optimization

The parameter ranges were selected to cover a practical spectrum of Random Forest complexity while keeping the search computationally manageable for the available dataset and experimental environment. The evaluated values were `n_estimators` = {10, 20, 30, 50, 75, 100} and `max_depth` = {3, 5, 10, 15, None}. The `n_estimators` values represent small to relatively large ensembles, allowing the study to observe the effect of increasing the number of trees on prediction stability and validation accuracy. The `max_depth` values represent shallow and moderate tree depths, while None allows unrestricted depth and provides a reference for a more complex model. Together, these ranges provide 30 candidate parameter combinations for systematic comparison using Stratified 5-Fold Cross-Validation.

Hyperparameter optimization was conducted using Stratified 5-Fold Cross-Validation to identify a Random Forest configuration that balances predictive accuracy and model complexity. The optimization focused on two key parameters: `n_estimators` and `max_depth`. Increasing `n_estimators` can improve ensemble stability and reduce variance at the expense of greater computational cost, whereas constraining `max_depth` limits tree complexity and can reduce the risk of overfitting. Systematic evaluation of all 30 candidate combinations allows the effect of these two parameters to be assessed jointly rather than optimizing only one parameter. The combination with the highest mean validation accuracy was selected for final model training.

2.5 Prediction and Evaluation (Confusion Matrix)

a Prediction Process

Following training across the candidate hyperparameter configurations, the prediction process evaluates the model's ability to classify unseen employee performance data. Within the Stratified 5-Fold Cross-Validation framework, the Random Forest model processes the validation fold in each iteration. Each decision tree independently classifies an input sample based on learned rules, and the final class label (e.g., contract renewal status) is determined by majority voting across the trees.

b Performance Evaluation Metrics

Model performance is evaluated using a confusion matrix within the cross-validation scheme to assess classification reliability and reduce data-splitting bias. To obtain stable estimates during hyperparameter selection, accuracy is computed for each validation fold and then averaged across the five folds. Overall accuracy measures the proportion of correctly classified employee records:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3}$$

For the positive class ("Renewed"), precision measures the proportion of true contract renewals among all positive predictions, thereby minimizing false-positive predictions:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

Conversely, recall measures the model's ability to correctly identify all employees whose contracts are actually renewed, thereby minimizing false-negative predictions:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5}$$

Finally, the F1-score provides a balanced assessment by calculating the harmonic mean of precision and recall, which is particularly useful when the dataset contains class imbalance:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

2.6 Evaluation Method for Test Results

After hyperparameter optimization was completed using only the training set, the selected Random Forest model was evaluated on the independent test set to measure its classification performance and provide an objective assessment of how closely its predictions matched the actual class labels.

Evaluation began with a confusion matrix, a two-dimensional table comparing predicted results with actual labels. Its four componentstrue positive (TP), true negative (TN), false positive (FP), and false negative (FN)represent the correct and incorrect predictions made by the model.

From these values, four evaluation metrics were calculated: accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correct predictions among all test records. Precision measures the accuracy of positive-class predictions, recall measures the model's ability to identify actual positive instances, and the F1-score is the harmonic mean of precision and recall. Together, these metrics provide a systematic assessment of Random Forest classification performance.

2.7 Summary of Research Methodology

The research began with the collection of quantitative data from the company's internal systems, covering employee attendance, individual competence, interpersonal competence, and operational performance from January to December 2024. All data were objective and anonymized to uphold research ethics.

The collected data then underwent preprocessing, including the selection of relevant attributes, standardization of values to a 1–4 Likert scale, and encoding of the target variable to distinguish employee contract renewal status. These steps ensured data quality and consistency and prepared the data for use with the Random Forest algorithm.

The Random Forest implementation focused on evaluating how two key parametersthe number of decision trees (`n_estimators`) and maximum tree depth (`max_depth`)affected model performance. Stratified 5-Fold Cross-Validation was applied during training to produce stable evaluation results and reduce the risk of overfitting.

3 RESULTS AND DISCUSSION

3.1 Data Collection

The data used in this study are secondary data obtained from the Human Resources (HR) department of PT Bringin Gigantara. The dataset consists of employee performance appraisal records used for classification with the Random Forest algorithm.

These data were used to predict the final performance appraisal outcome in two categories: "Renewed" and "Not Renewed." The dataset consists of 10 independent variables (assessment criteria) and one dependent target variable.

3.2 Data Preprocessing Stage

Data preprocessing is a crucial step performed prior to applying the Random Forest algorithm. Its main purpose is to ensure that the data used is of good quality, consistent, and suitable for the requirements of the machine learning model, thereby enabling optimal classification performance.

a Data Selection

In the data selection stage, 100 out of 132 available employees were selected, as the remaining 32 held managerial or permanent positions and were excluded, since the analysis focused solely on contract employees. The Employee ID and Name attributes were used as identifiers, while variables V1–V10 were selected as the main performance assessment attributes and prepared for the preprocessing stage.

b Data Cleaning

During data cleaning, the dataset was checked for completeness and consistency. No missing values, duplicate entries, or out-of-range values were found; therefore, no additional cleaning was required, and the data were ready for analysis. The data-cleaning process is illustrated in Table 1.

Table 1. Data Cleaning

NIK	Nama Karyawan	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
001	Karyawan_001	18	36	45	30	10	43	69	33	43	9
002	Karyawan_002	63	63	63	60	70	50	65	69	69	55
003	Karyawan_003	47	24	18	70	40	50	50	60	21	41
004	Karyawan_004	63	53	63	50	60	50	75	67	63	73
005	Karyawan_005	73	100	67	40	100	50	74	91	54	82
006	Karyawan_006	17	67	33	40	60	20	63	36	9	27
007	Karyawan_007	59	94	59	40	80	17	41	25	25	45
008	Karyawan_008	11	44	11	40	30	20	49	40	13	18
009	Karyawan_009	73	64	45	80	50	33	63	53	76	68
010	Karyawan_010	50	86	43	70	60	57	68	52	71	50
...	...	...	...	...	...	...	...	...	...	...	...
100	Karyawan_100	65	82	53	40	80	67	64	98	65	64

c Encoding

After data cleaning, the employee data were transformed into a more structured format. All performance-variable values were rescaled to a 1–4 Likert scale to standardize the range across variables. The resulting variables were labeled X1–X10.

The target variable representing contract renewal status was encoded numerically, with 1 for the "Renewed" category and 0 for the "Not Renewed" category. The encoded data are shown in Table 2.

Table 2. Data Encoding

NIK	K	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y
001	K1	2	2	3	2	3	3	3	2	3	2	1
002	K2	3	3	3	3	3	3	3	3	3	3	1
003	K3	2	1	2	3	2	3	3	3	2	2	0
004	K4	3	3	3	3	3	3	4	3	3	3	1
005	K5	2	3	2	2	3	3	3	2	3	2	1
006	K6	2	3	2	2	3	2	3	2	2	2	0
007	K7	3	4	3	2	4	2	2	2	2	2	1
008	K8	2	2	2	2	2	2	2	2	2	2	0
009	K9	3	3	2	4	3	2	3	3	4	3	1
010	K10	3	4	3	3	3	3	3	3	3	3	1
...	...	...	...	...	...	...	...	...	...	...	...	...
100	K100	3	4	3	2	4	3	3	4	3	3	1

d Data Split

The dataset was divided into 80% training data (80 employee records) and 20% test data (20 employee records). The training set consisted of 70 Renewed and 10 Not Renewed records, while the test set consisted of 15 Renewed and 5 Not Renewed records. Hyperparameter optimization was performed exclusively on the training

set using Stratified 5-Fold Cross-Validation with `n_splits` = 5, `shuffle`=True, and `random_state` = 42. The independent test set was reserved solely for final evaluation and was not used to select the hyperparameter configuration.

3.3 Random Forest Implementation

At this stage, the Random Forest algorithm was applied to classify the performance of PT Bringin Gigantara employees based on the preprocessed evaluation data. The goal was to build a classification model capable of objectively identifying patterns in the relationship between performance assessment variables and the final employee evaluation outcome.

a Cross Validation and Parameter Optimization

Stratified 5-Fold Cross-Validation was used to search for and select the best hyperparameter configuration, with each candidate combination evaluated across five stratified data splits. The mean validation accuracy across all five folds served as the basis for selecting the optimal configuration.

In the parameter optimization stage, the key model parameters affecting classification complexity and performan`n_estimators` (number of trees) and `max_depth` (maximum tree depth)were tuned jointly to balance accuracy and generalization ability. The 30 combinations formed by `n_estimators` = {10, 20, 30, 50, 75, 100} and `max_depth` = {3, 5, 10, 15, None} were evaluated using Stratified 5-Fold Cross-Validation on the training set.

This configuration was subsequently used as the final model for the prediction and evaluation stages. It was selected to provide strong accuracy while maintaining balanced generalization performance on previously unseen data. The comparative results of all parameter combinations are presented in Table 3.

Table 3. Parameter Optimization

n_estimators	max_depth	Fold1	Fold2	Fold3	Fold4	Fold5	CV Mean
10	None	0.80	1.00	0.95	1.00	0.95	0.94
10	3	0.75	1.00	1.00	1.00	0.95	0.94
10	5	0.80	1.00	0.95	1.00	0.95	0.94
10	10	0.80	1.00	0.95	1.00	0.95	0.94
10	15	0.80	1.00	0.95	1.00	0.95	0.94
20	None	0.75	1.00	1.00	1.00	0.95	0.94
20	3	0.80	0.95	1.00	1.00	0.95	0.94
20	5	0.75	1.00	1.00	1.00	0.95	0.94
20	10	0.75	1.00	1.00	1.00	0.95	0.94
20	15	0.75	1.00	1.00	1.00	0.95	0.94
30	None	0.85	1.00	1.00	1.00	0.95	0.96
30	3	0.80	0.95	1.00	1.00	0.95	0.94
30	5	0.85	1.00	1.00	1.00	0.95	0.96
30	10	0.85	1.00	1.00	1.00	0.95	0.96
30	15	0.85	1.00	1.00	1.00	0.95	0.96
50	None	0.80	1.00	1.00	1.00	0.95	0.95
50	3	0.75	0.95	1.00	1.00	0.95	0.93
50	5	0.80	1.00	1.00	1.00	0.95	0.95
50	**10**	**0.95**	**1.00**	**1.00**	**1.00**	**0.95**	**0.98**
50	15	0.95	1.00	1.00	1.00	0.95	0.98
75	None	0.95	1.00	1.00	1.00	0.95	0.98
75	3	0.80	0.95	1.00	1.00	0.95	0.94
75	5	0.80	1.00	1.00	1.00	0.95	0.95
75	10	0.95	1.00	1.00	1.00	0.95	0.98
75	15	0.95	1.00	1.00	1.00	0.95	0.98
100	None	0.95	1.00	1.00	1.00	0.95	0.98
100	3	0.80	0.95	1.00	1.00	0.95	0.94
100	5	0.95	1.00	1.00	1.00	0.95	0.98
100	10	0.95	1.00	1.00	1.00	0.95	0.98
100	15	0.95	1.00	1.00	1.00	0.95	0.98

To improve result transparency, fold-level accuracies are reported instead of presenting only the average value. The selected hyperparameter combination was determined based on the highest mean validation accuracy together with the stability of the fold results.

3.4 Model Performance Evaluation

After the best parameter combination was obtained, the next step was model performance evaluation. The Random Forest model was trained using the selected parameters and then tested to measure its performance in classifying employee performance. This evaluation assessed the model's accuracy and generalization ability on previously unseen data.

a Confusion Matrix – Training Data

Model performance on the training set was evaluated using a confusion matrix comparing predicted results with actual labels. Figure 2 shows the confusion matrix for the training set. The model correctly classified 68 instances of class 1 and 9 instances of class 0, with one misclassification in class 0 and no errors in class 1. This resulted in an accuracy of 98.75%, indicating strong performance on the training set.

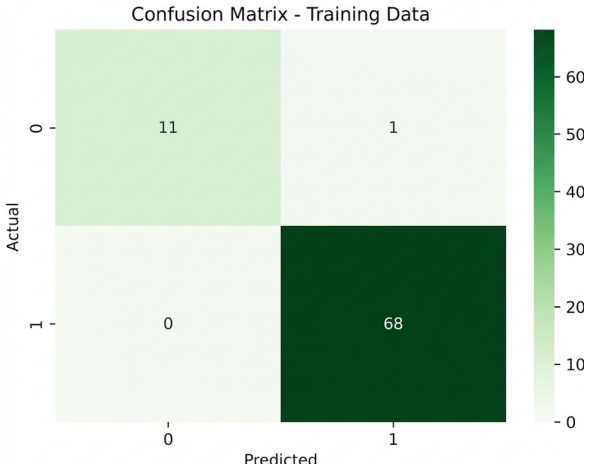

Figure 2. Training Data Confusion Matrix

Based on the confusion matrix, the model achieved an accuracy of 98.75%. Macro precision was 99.18%, while macro recall was 96%. The macro F1-score was 97%, indicating a strong balance between precision and recall across the two classes, as depicted in Figure 3. Overall, these results indicate that the classification model performed well on the training set and can support employee contract renewal decisions.

$$Accuracy = \frac{68+11}{68+11+0+1} = \frac{79}{80} = 98{,}75\%$$

$$Precision_{macro} = \frac{Preci_0 + Preci_1}{2} = \frac{1.00+0.99}{2} = 99.18\%$$

$$Recall_{macro} = \frac{Recall_0 + Recall_1}{2} = \frac{0.92+1.00}{2} = \frac{1.92}{2} = 96\%$$

$$F1_{macro} = \frac{F1_0 + F1_1}{2} = \frac{0.96+0.99}{2} = 97\%$$

Figure 3. Confusion Matrix Calculation Results for Training Data

b Confusion Matrix – Testing Data

After evaluation on the training set, the model was tested on the test set to assess its performance on previously unseen data and provide a more objective measure of generalization. Figure 4 shows the confusion matrix for the test set. The model correctly classified 17 instances of class 1 and 2 instances of class 0, with one misclassification in class 0. This resulted in an accuracy of 95%, indicating good generalization performance on the test set.

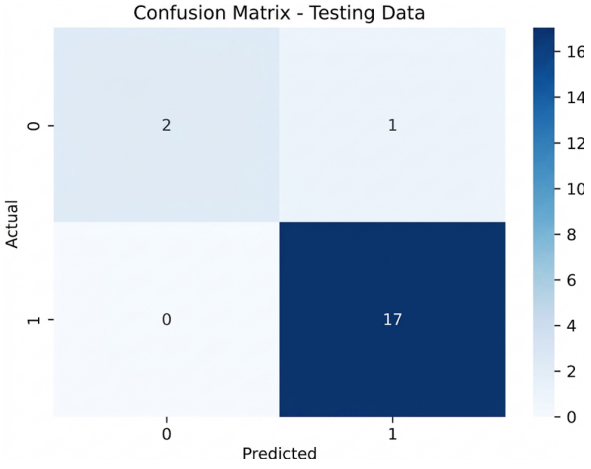

Figure 4. Testing Data Confusion Matrix

Based on the test results, the model achieved an accuracy of 95%, indicating that most records were correctly classified. Macro precision, macro recall, and macro F1-score were each 97%, as calculated in Figure 5. These results demonstrate that the model classifies the test records effectively and generalizes adequately to unseen data.

$$Accuracy = \frac{17+2}{17+2+1+0} = \frac{19}{20} = 95\%$$

$$Precision_{macro} = \frac{Preci_{0+}Preci_{1}}{2} = \frac{1.00+94}{2} = 97\%$$

$$Recall_{macro} = \frac{Recall_{0+}Recall_{1}}{2} = \frac{67+1.00}{2} = 83\%$$

$$F1_{macro} = \frac{F1_{0+}F1_{1}}{2} = \frac{80+97}{2} = 89\%$$

Figure 5. Confusion Matrix Calculation Results for Testing Data

Table 4 summarizes the model's performance on the training and test sets. Performance on the training set was higher than on the test set, which is expected because the model was fitted to the training data. The results nevertheless indicate reasonably good generalization.

Table 4. Comparison of Training and Test Sets

Evaluation Metric	Training Data (80%)	Testing Data (20%)
Accuracy (%)	98	95
Precision (%)	99	97
Recall (%)	96	83
F1-Score (%)	97	89

The test set is relatively small, comprising only 20 employee records (15 Renewed and 5 Not Renewed), which may limit the statistical robustness of the reported test accuracy. To provide additional evidence of model reliability, the Stratified 5-Fold Cross-Validation results in Table 3 offer a more comprehensive validation. The selected configuration (`n_estimators` = 50, `max_depth` = 10) achieved a mean cross-validation accuracy of 0.98 across the folds during training, indicating stable performance beyond a single train-test split. Future research should validate the model using a larger and more balanced employee dataset to strengthen the generalizability of the findings.

Although the independent test set contains only 20 employee records, model validation was not based solely on this hold-out set. Stratified 5-Fold Cross-Validation was performed on the training set during hyperparameter optimization; therefore, the reported test accuracy should be interpreted together with the cross-validation results presented in Table 3.

c Employee Prediction Results

In addition to the confusion matrix evaluation, this study presents the model's prediction results for the employee data. Table 5 compares the actual labels with the Random Forest predictions and shows that the model successfully classified employee performance status into the "Renewed" and "Not Renewed" categories. Most predictions matched the actual labels, indicating that the model captured the relationships among the performance evaluation variables effectively.

Overall, the high prediction accuracy indicates that the selected hyperparameters produced strong model performance, consistent with the test results. Accordingly, these prediction results can serve as supporting information for employee performance evaluation and contract renewal decisions.

Table 5. Employee Prediction Results

NIK	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Actual	Prediction
37	3	4	3	2	4	4	3	4	4	3	1	1
45	3	3	2	3	3	2	3	2	4	2	1	1
4	3	3	3	3	3	3	4	3	3	3	1	1
2	3	3	3	3	3	3	3	3	3	3	1	1
27	2	3	3	3	3	3	2	3	3	2	1	1
73	3	4	2	3	3	2	2	3	2	2	1	1
49	3	4	3	2	4	4	3	4	4	3	1	1
11	3	2	2	3	3	2	2	1	2	3	0	1
60	4	3	4	2	3	4	4	3	4	4	1	1

3.5 Result Analysis

The optimized Random Forest model achieved a test accuracy of 95%. The parameter search results show that `n_estimators` = 50 and `max_depth` = 10 achieved a mean Stratified 5-Fold Cross-Validation accuracy of

0.98. This configuration was selected because it reached the highest validation performance while maintaining a relatively simple ensemble compared with configurations using more trees or deeper trees. The results also show that increasing `max_depth` from 5 to 10 improved the mean validation accuracy for `n_estimators` = 50 from 0.95 to 0.98, indicating that shallow trees were not sufficiently expressive for the observed employee performance patterns. Increasing `max_depth` beyond 10 did not improve the mean accuracy, while increasing `n_estimators` from 50 to 75 or 100 at `max_depth` = 10 also produced 0.98. Thus, `n_estimators` = 50 and `max_depth` = 10 provide a practical balance between predictive performance, model complexity, and computational cost rather than simply maximizing the number or depth of trees.

Table 6. Comparison with Related Employee-Performance Studies

Study	Method	Accuracy
Previous employee classification study	Random Forest (default parameters)	88–93%
Career placement classification	Decision Tree / Naïve Bayes	85–91%
This study	**Optimized Random Forest**	**95%**

The 95% test accuracy is comparable to the 95.83% Random Forest accuracy reported by Patel et al. [12] for employee-performance classification, although the datasets, target definitions, and experimental settings are different and therefore the values should not be interpreted as a direct head-to-head comparison. In another employee-performance prediction study, Singh and Khuteja reported 87% accuracy using Random Forest, whereas the present model achieved 95% on the independent test set. These comparisons indicate that the proposed model performs competitively with existing Random Forest-based employee-performance studies. More importantly, the contribution of the present study is methodological: rather than relying on a default Random Forest configuration, it jointly evaluates `n_estimators` and `max_depth` using Stratified 5-Fold Cross-Validation and selects a configuration that maintains strong validation performance with a comparatively compact ensemble. Tanasescu et al. [4] likewise demonstrate the value of preprocessing, model comparison, and hyperparameter tuning for employee-performance prediction; the present study differs by focusing specifically on joint optimization of these two Random Forest parameters for employee contract-renewal classification.

4 CONCLUSION

This study demonstrates that systematic hyperparameter optimization using Stratified 5-Fold Cross-Validation can provide a stable Random Forest configuration for employee performance classification. The selected configuration, `n_estimators` = 50 and `max_depth` = 10, achieved a mean cross-validation accuracy of 98% and a test accuracy of 95%, indicating good predictive performance on the independent test set. The practical implication is that the model can be used as a decision-support tool by human resource personnel to organize performance evidence, identify employees predicted as "Renewed" or "Not Renewed," and support more objective, transparent, and data-driven contract renewal discussions. The model should support rather than replace managerial judgment and established HR policies. A key limitation is the relatively small dataset of 100 contract employees from a single company, with only 20 records in the independent test set; this limits the statistical robustness and generalizability of the findings to other organizations. Future research should therefore validate the model on larger and more diverse multi-company datasets, consider more balanced class distributions, and evaluate additional Random Forest hyperparameters or alternative ensemble optimization methods.

ACKNOWLEDGMENTS

The authors would like to thank PT Bringin Gigantara and Universitas Indo Global Mandiri for supporting this research.

DECLARATIONS

Author Contributions:

M. Imam Saputra: Conceptualization, Methodology, Software, Writing–original draft.
Rendra Gustriansyah: Supervision, Methodology, Writing–review & editing.
Dwi Asa Verano: Data curation, Validation, Formal analysis.

Funding:

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflicts of Interest:

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability:
The data that support the findings of this study are available from the corresponding author upon reasonable request.

AI Use Statement:
During the preparation of this work, the authors used Qwen AI in order to improve the language and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the final content of the published article.

Additional Information:
No additional information is available for this paper.

REFERENCES

[1] S. Amaliah and M. Nusrang, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng," *J. Variansi*, vol. 4, no. 2, pp. 121–127, 2022, doi: 10.35580/variansi-unm31.

[2] E. Y. Boateng, J. Otoo, and D. A. Abaye, "Basic Tenets of Classification Algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A Review," *J. Data Anal. Inf. Process.*, vol. 8, no. 4, pp. 341–357, 2020, doi: 10.4236/jdaip.2020.84020.

[3] G. G. Casas and Z. H. Ismail, "Computer Vision, Machine Learning, and Deep Learning for Wood and Timber Products: A Scopus-Based Bibliometric and Systematic Mapping Review (1983–2026)," *IEEE Access*, pp. 1–29, 2026.

[4] L. G. Tanasescu, A. Vines, and A. R. Bologa, "Data Analytics for Optimizing and Predicting Employee Performance," *Appl. Sci.*, vol. 14, no. 12, p. 5120, 2024.

[5] G. Technica, "Multi-Algorithm Classification and CA–Markov Prediction of Land-Use Change Across Three Temporal Intervals: A Case Study of the Loe Tor Royal Project, Tak Province, Thailand," *Geogr. Tech.*, vol. 21, pp. 206–229, 2026, doi: 10.21163/GT.

[6] N. S. Thomas and S. Kaliraj, "An Improved and Optimized Random Forest Based Approach to Predict the Software Faults," *SN Comput. Sci.*, vol. 5, no. 5, p. 512, 2024, doi: 10.1007/s42979-024-02764-x.

[7] R. Bagus, I. Kusuma, and G. Prajitno, "Analisis dan Simulasi Optimasi Parameter Akustik Ruang pada Smart Classroom," *J. Tek. ITS*, vol. 10, no. 2, pp. 112–118, 2021.

[8] L. Qian, Q. Zuo, D. Li, and H. Zhu, "DGMSCL: A dynamic graph mixed supervised contrastive learning approach for class imbalanced multivariate time series classification," *Neural Networks*, vol. 185, p. 107131, 2025, doi: 10.1016/j.neunet.2025.107131.

[9] R. G. Guntara, "Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 1, pp. 55–60, 2023.

[10] G. M. Agung, R. A. Zuama, and E. S. Budi, "Analysis of Student Academic Performance Using Random Forest and Support Vector Machines," *J. Inf. Syst. Explor.*, vol. 6, no. 1, pp. 57–65, 2026.

[11] W. Shuwei and E. Law, "Environmental sustainability: a major component of sustainable development," *Environ. Sci. Pollut. Res.*, vol. 30, pp. 900–907, 2023.

[12] K. Patel et al., "RanKer: An AI-Based Employee-Performance Classification Scheme to Rank and Identify Low Performers," *IEEE Access*, vol. 10, pp. 1–21, 2022.

[13] L. Lazar and E. Ristea, "Smart Solutions for Mitigating Eutrophication in the Romanian Black Sea Coastal Waters Through an Integrated Approach Using Random Forest, Remote Sensing, and System Dynamics," *Environ. Monit. Assess.*, vol. 198, pp. 1–26, 2026.

[14] A. Smith, "Random Forest Algorithm for Classification of Multiwavelength Data," *Astron. Astrophys.*, vol. 500, pp. 120–130, 2009.

[15] I. Tasya, M. Raihan, L. Luren, and W. Sapitri, "Analisis Pengaruh Arus Globalisasi Terhadap Budaya Lokal di Era Digital," *J. Hum. Sos.*, vol. 3, no. 2, pp. 40–47, 2023.

[16] M. R. Adrian, M. P. Putra, M. H. Rafialdy, N. A. Rakhmawati, and D. S. Informasi, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *J. Res. Comput. Sci.*, vol. 7, no. 1, pp. 36–40, 2021.

[17] A. Rajamanickam and C. Kamalakannan, "Land Use and Land Cover Prediction in Tamilnadu of India, Using Random Forest Machine Learning Technique," *J. Indian Soc. Remote Sens.*, vol. 20, pp. 145–155, 2025.

[18] O. Pahlevi and Y. Handrianto, "Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit," *J. Comput. Sci.*, vol. 5, no. 1, pp. 71–76, 2023.

[19] C. Kopitsa, I. G. Tsoulos, V. Charilogis, and A. Stavrakoudis, "Predicting the Duration of Forest Fires Using Machine Learning Methods," *Fire*, vol. 7, no. 3, pp. 1–19, 2024.

[20] W. C. Wahyudin, "A Cluster-Label-Based Framework for Water Quality Risk Pattern Classification Using Naive Bayes and Random Forest," *J. Environ. Inform.*, vol. 10, no. 3, pp. 1511–1522, 2026.