JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 9, No. 1, June 2026, pp. 425~436
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v9i1.5639

Application of Backward Elimination Method for Optimization of Decision Tree C4.5 Algorithm in Employee Performance Prediction

Ahmad Fauzi[1], Novita Indriyani[2], Andika Bayu Hasta Yanto[3]
[1,3] Teknik dan Informatika, Universitas Bina Sarana Informatika, Jl.Kramat Raya No.98, Senen, Jakarta Pusat
[2]Teknik dan Informatika, Universitas Bina Sarana Informatika Kampus Kota Bogor, Komplek Pesona Intiland Jl. Raya Cilebut Kel. Sukaresmi Tanah Sareal - Bogor

Article Info

Article history:

Received 05 01, 2026
Revised 06 11, 2026
Accepted 06 25, 2026

Keywords:

Employee Performance Prediction
Decision Tree C4.5
Backward Elimination
Feature Selection
Classification Accuracy

ABSTRACT

This study aims to analyze the application of the Backward Elimination method in optimizing the Decision Tree C4.5 algorithm in predicting employee performance. The main problem in this study is the employee performance evaluation process which is still done manually so that it has the potential to cause inconsistency in the assessment results. The study used a dataset of 500 employee data with several performance assessment attributes such as age, education level, work discipline, productivity, and superior assessment. The research method includes data preprocessing, feature selection using Backward Elimination, application of the Decision Tree C4.5 algorithm, and model evaluation using 10-Fold Cross Validation in the RapidMiner application. The test results show that the Decision Tree C4.5 algorithm without optimization obtained an accuracy value of 89.60% and an AUC of 0.944. After applying the Backward Elimination method, model performance increased with an accuracy value of 92.80% and an AUC of 0.972. This increase indicates that the Backward Elimination method is able to reduce less relevant attributes so that the classification process becomes more optimal. Thus, the application of the Backward Elimination method has proven effective in improving the performance of the Decision Tree C4.5 algorithm in predicting employee performance.

Corresponding Author:

Ahmad Fauzi
Department of Engineering and Informatics
Universitas Bina Sarana Informatika
Jakarta, Indonesia
Email: ahmad.afz@bsi.ac.id

1. Introduction

Employee performance is one of the primary factors that determine an organization's success in achieving business objectives and delivering high-quality public services. Employees with strong work performance contribute significantly to organizational productivity, efficiency, and competitiveness, while employees with poor performance can negatively affect operational quality and overall organizational effectiveness. Therefore, organizations need to conduct periodic performance evaluations to measure employee productivity and identify areas for improvement.

In practice, employee performance evaluation is often conducted manually based on subjective assessments from supervisors. This process requires considerable time and effort and may result in inconsistencies due to differences in evaluators' perceptions and judgments. Furthermore, as the number of employees increases, the evaluation process becomes more complex, making it difficult for organizations to maintain accuracy and consistency in performance assessments.

The rapid development of data mining and machine learning technologies provides opportunities for organizations to analyze large volumes of employee data more efficiently and accurately. Data mining techniques enable the extraction of meaningful patterns from historical data, which can support decision-making processes and improve performance management systems [1], [2]. Among various classification algorithms, the Decision Tree C4.5 algorithm is widely used because of its ability to generate classification rules that are easy to interpret and implement. The algorithm constructs a decision tree by selecting attributes with the highest information gain, allowing it to classify data effectively and transparently [3].

However, the presence of numerous attributes in a dataset may reduce classification performance due to irrelevant or redundant features. These unnecessary attributes can increase model complexity, lead to overfitting, and decrease prediction accuracy [4]. Consequently, feature selection techniques are required to identify the most influential attributes and eliminate less relevant variables. One of the feature selection methods commonly used is Backward Elimination. This method systematically removes attributes that contribute the least to the predictive model, thereby improving classification efficiency and accuracy. Feature selection methods have been widely recognized as effective approaches for improving classification performance and reducing computational complexity [5], [6].

Several previous studies have demonstrated that optimization techniques can enhance the performance of classification algorithms. For example, the application of Particle Swarm Optimization (PSO) to Decision Tree algorithms has been shown to improve classification accuracy and model effectiveness [7]. Similarly, the implementation of feature selection techniques such as Forward Selection on the C4.5 algorithm has produced better prediction results by selecting the most relevant features. These findings indicate that optimization and feature selection methods can significantly improve the effectiveness of classification models in various domains.

Based on these considerations, this study aims to compare the performance of the standard Decision Tree C4.5 algorithm and the Decision Tree C4.5 algorithm optimized using the Backward Elimination method in predicting employee performance. The proposed approach is expected to improve classification accuracy by eliminating irrelevant attributes and identifying the most influential factors affecting employee performance. The results of this research are expected to provide valuable insights into the effectiveness of feature selection techniques in enhancing classification performance and supporting more objective, accurate, and data-driven employee evaluation processes.

2. Research Method

This study was conducted to analyze the application of the Backward Elimination method in optimizing the Decision Tree C4.5 algorithm for employee performance prediction. The research methodology consisted of several stages, including data collection, data preprocessing, classification model development, feature selection, and model performance evaluation. These stages were carried out systematically to ensure the validity and reliability of the research results.

The data collection stage involved gathering employee performance data from organizational records. The dataset contained several attributes related to employee performance, such as age, education level, work discipline, productivity, and supervisor assessment. After the data were collected, a preprocessing stage was performed to improve data quality. This process included data cleaning, handling missing values, data transformation, and data integration to ensure that the dataset was suitable for the classification process [8], [9].

The next stage involved developing a classification model using the Decision Tree C4.5 algorithm. Decision Tree C4.5 is a widely used classification technique that generates decision rules based on information gain values and produces interpretable classification models [10]. To improve model performance, the Backward Elimination method was applied as a feature selection technique. This method removes attributes that contribute the least to the classification model, thereby reducing irrelevant features and improving prediction accuracy [11], [12].

Model evaluation was conducted using 10-Fold Cross Validation to obtain reliable performance measurements and minimize bias in the testing process [13]. The performance of the standard Decision Tree C4.5 model and the optimized model with Backward Elimination was compared using accuracy and Area Under Curve (AUC) metrics. The entire experimental process was implemented using the RapidMiner application. The research stages were carried out systematically as illustrated in the research flowchart shown in the following figure.

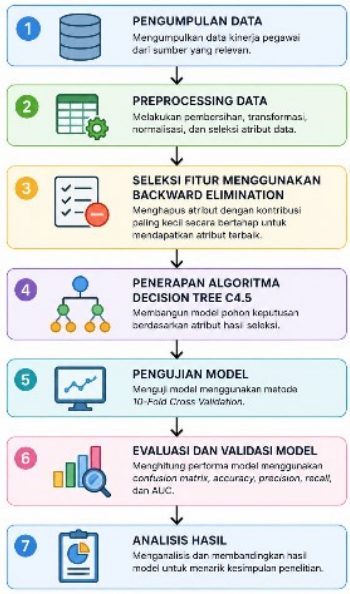

Figure 1. Research Stages

This research was conducted systematically to determine the effect of the Backward Elimination method in improving the performance of the Decision Tree C4.5 algorithm in predicting employee performance. The research stages began with data collection and analysis of the classification model test results.

1. Data Collection

The initial stage of the research involved collecting employee performance data obtained from company evaluations. The dataset used consisted of several attributes related to employee performance assessments, such as age, education level, work discipline, absence, productivity, communication skills, and superior assessments. This data served as the basis for the employee performance classification process.

The data collection stage is crucial because data quality will affect the classification results generated by the machine learning model. The better the quality of the data used, the better the model's performance.

2. Data Preprocessing

The data preprocessing stage was conducted to improve the quality of the dataset before the classification process. This stage includes data cleaning, data transformation, data normalization, and the removal of incomplete or duplicate data. Preprocessing aims to make the data more structured and ready for processing by the classification algorithm. In addition, preprocessing also helps reduce noise in the data so that the classification results are more optimal.

3. Feature Selection Using Backward Elimination

The next stage is the feature selection process using the Backward Elimination method. This method is used to remove attributes that have the least contribution to the classification model. The Backward Elimination process is carried out in stages by evaluating each attribute used in the dataset. Attributes deemed less relevant are removed to obtain the best combination of attributes that can improve the performance of the Decision Tree C4.5 algorithm.

The use of this feature selection method aims to increase model efficiency, reduce data complexity, and improve classification accuracy.

4. Implementation of the Decision Tree C4.5 Algorithm

After the feature selection process is complete, the next stage is the application of the Decision Tree C4.5 algorithm to build a classification model for predicting employee performance. The C4.5 algorithm works by forming a decision tree based on the highest gain value for each attribute. This process is repeated until all data is successfully classified into a specific class. The Decision Tree model was chosen because it is able to produce classification rules that are easy to understand and has a fairly good level of accuracy in the data classification process.

5. Model Testing

The model testing stage was carried out using the 10-Fold Cross Validation method in the RapidMiner application. In this method, the dataset is divided into 10 parts, then training and testing are performed alternately until all data is used as testing data. Cross-validation is used to determine the level of model stability and reduce the risk of overfitting in the classification process.

6. Model Evaluation and Validation

The evaluation stage is conducted to determine the performance of the resulting classification model. Model evaluation is performed using a confusion matrix and ROC Curve to obtain accuracy, precision, recall, and Area Under Curve (AUC) values. The accuracy value is used to determine the level of classification accuracy, while the AUC is used to measure the model's ability to distinguish data classes. The higher the accuracy and AUC values, the better the performance of the resulting classification model.

7. Results Analysis

The final stage is the analysis of the model testing results. In this stage, the performance of the Decision Tree C4.5 algorithm is compared without optimization and the Decision Tree C4.5 algorithm optimized using the Backward Elimination method. The analysis results are used to determine the effect of the Backward Elimination method on improving the accuracy and performance of the classification model in predicting employee performance.

C4.5 Decision Tree Algorithm

The C4.5 algorithm is a classification method used to create a predictive model based on data organized into a decision tree[14].

The steps of the C4.5 algorithm include:

1. Calculating the entropy value
2. Calculating the information gain value
3. Determining the attribute with the highest gain as the root node
4. Forming the decision tree branches
5. Repeating the process until all data is classified

The entropy calculation is performed using the following equation[15]:

$$Enthropy(S) = -\sum_{i=1}^{i} \frac{S_i}{S} log_2 \frac{S_i}{S} \tag{1}$$

Sedangkan perhitungan information gain dilakukan menggunakan persamaan[2]:

$$Gain(S,A) = Enthropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Enthropy(Si) \tag{2}$$

Previous Research

Based on research in the journal "Implementation of the C4.5 Algorithm with the Backward Elimination Feature Selection for MSME Product Sales Strategy," it was concluded that: Based on the research results, data was collected on MSMEs in East Kedungwuni District, Pekalongan Regency, Central Java. The data was then extracted and processed using data mining techniques. The algorithm used was the C4.5 algorithm, which produced an accuracy of 82.78%. The addition of the backward elimination feature resulted in an accuracy of 85.00%. The addition of the backward elimination feature was proven to increase accuracy compared to using the C4.5 algorithm alone. [3]

Based on research in the journal "Optimization of the C4.5 Algorithm Using the Forward Selection and Stratified Sampling Methods for Creditworthiness Prediction," it was concluded that: The C4.5 algorithm proved effective in predicting creditworthiness with an accuracy rate of 79.11%. The Forward Selection and Stratified Sampling methods have been proven to increase the accuracy of the C4.5 algorithm by 9.2% in predicting creditworthiness [4].

Based on research in the journal "Predicting Student Graduation Using the Decision Tree C4.5 Algorithm with Pruning Techniques," it was concluded that: Experiments comparing the accuracy of selected tree models based on the number of folds in the k-fold cross-validation process showed that the higher the fold value, the greater the accuracy of the selected tree model [5].

Based on research in the journal "Comparison of the Application of the Decision Tree C.45 Algorithm and Naïve Bayes in Analyzing Student Graduation at SMK Swadhipa 2 Natar, South Lampung Regency," it was concluded that: Decision Tree C4.5 has higher accuracy, precision, and F1-score than Naïve Bayes. However, Naïve Bayes has slightly higher recall [6].

Based on research in the journal "The Utilization of Decision Tree Algorithm in Order to Predict Heart Disease," it concludes that: The results of this study can be considered by experts to assist decision-making in treating heart disease. Furthermore, for further research, it is recommended to create dashboards and visualizations of the relationships between features that influence heart disease [7].

Based on research in the journal "Comparison of the Performance of the C.45 Algorithm with Naive Bayes in Analyzing Book Borrowing at the Library of Pringsewu Muhammadiyah University," it concludes that: The results show that the C4.5 algorithm outperforms Naive Bayes in all evaluation metrics, including accuracy, precision, recall, and F-score. Specifically, C4.5 achieved an accuracy of 96.26%, while Naive Bayes achieved 91.44%. This indicates that C4.5 is more effective in capturing complex relationships in data and predicting user borrowing behavior. The findings of this study underscore the importance of using data mining methodologies in library management. By implementing the C4.5 algorithm, the Pringsewu Muhammadiyah University Library can improve its services, such as personalized book recommendations, targeted marketing efforts, and more efficient budgeting for new acquisitions[8].

Based on research in the journal entitled "Implementation of the C4.5 Algorithm Through a Decision Tree Based on the Forward Selection Method to Predict Bad Credit Risk," it concluded that: This processing tested two different scenarios: the basic C4.5 algorithm and the C4.5 algorithm that had been enhanced with the forward selection method. The C4.5 test results without enhancements obtained an average accuracy of 83.33%, with a class recall of 88.24% true smooth, and 76.02% true risk of bad credit. The C4.5 test results with enhancements obtained an average accuracy of 87.59%, categorized as good classification[9].

Based on research in the journal entitled "Implementation of the C4.5 Classification Algorithm to Predict Vehicle Purchase Feasibility" concluded that: The methodology used is a literature study methodology to select the best algorithm to process the existing dataset. The C4.5 algorithm is the algorithm used selected in this study. The C4.5 algorithm is then implemented to determine the car that will be taken by the buyer according to the background criteria of buying (purchase price), maint (maintenance costs), doors (number of doors), persons (capacity to accommodate passengers), lug_boot (trunk area), safety (vehicle safety level). The results of the experiment are classified as good where the results shown are 92.4% accuracy, 92.4% Precision and 92.4% Recall [10].

Based on research in the journal entitled "Application of the C4.5 Algorithm to Clarify Social Assistance Receipts Using Feature Selection" concluded that: Thus, this research was obtained in the research on KIP receipts at the Sesai base above in this study, it provides convenience in improving the data verification process to distribute KIP assistance in the Sesai base area. The attribute that most influences it is A1 or ART from 1675 initial data tested using a comparison of 70% test data and 30% training data, an accuracy of 98.21% was obtained, a precision value of 98.18% and a recall value of 93.91% thus the results of the C4.5 algorithm can work very well in determining KIP receipts at the Sesai base, with an accuracy of 98.21% from here we can conclude that the use of the C4.5 algorithm has very good accuracy, if it can also be compared with other algorithms in order to produce results that may be even better [11]. Based on research in the journal "Implementation of the Decision Tree Algorithm for Classifying Best-Selling Products," it was concluded that: The accuracy of the best-selling product classification model using the Decision Tree C4.5 obtained from this study was 90% and an AUC value of 0.709, which is included in the Good Classification category. Therefore, it can be concluded that the Decision Tree C4.5 data mining classification model is accurate in classifying best-selling products [12].

Based on research in the journal "C4.5 Algorithm for Determining Credit Eligibility (Case Study: Bank Mandiri Taspen Sukabumi Cash Office)," it was concluded that: The results of the study, which used 258 private customer data from November and December 2021 at the Bank Mandiri Taspen Sukabumi Cash Office, resulted in an evaluation that the C4.5 Algorithm was accurately applied to predict whether or not customer credit card payments were in default, with an accuracy level of 93.75% for training data of 0.9 and testing data of 0.1. Furthermore, an accuracy level of 96.77% was obtained for training data of 0.8 and testing data of 0.2 [13].

Based on research in the journal entitled "Implementation of C4.5 Data Mining in Measuring the Level of Student Satisfaction with Academic Services," it concludes that: The implementation in this study resulted in a decision on student satisfaction with academic services in the Informatics Study Program, Faculty of Computer Science, UPS Tegal, obtaining a model, rule, and prediction of student satisfaction with an accuracy score of 87.95% and an AUC score of 0.995, thus entering the very good data category. This is because tangible attributes and empathy significantly influence the level of student satisfaction [14].

Based on research in the journal entitled "Decision Tree Using the C4.5 Algorithm for Credit Feasibility Analysis," it concludes that: The C4.5 algorithm with a decision tree can be used to analyze the creditworthiness of prospective debtors. The rules generated by the decision tree can be used as a basis for making credit policies at PT BPR Lubuk Raya Mandiri. The results of the resulting decision tree are not only in the form of a decision of whether it is feasible or not feasible. But also used to create various credit granting policies based on the resulting pattern [15].

Based on research in the journal entitled "Comparison of C4.5 and Naïve Bayes Classification Methods to Measure Customer Satisfaction" concluded that: 1. Comparison of Data mining classification methods, C4.5 and Naïve Bayes, shows that C4.5 is more accurate than the Naïve Bayes method. This is proven by the accuracy value where the C4.5 method has an accuracy value of 94.17%, while Naïve Bayes 85.83%. 2. Based on the C4.5 method, the Assurance attribute is the attribute that most influences company performance. Viewed from the Assurance attribute as the root node. 3. Based on the AUC value, Naïve Bayes is included in the very good category, C4.5 is included in the poor classification category [16].

Based on research in the journal entitled "Determination of Factors Affecting Customer Satisfaction Using the C4.5 Algorithm" concluded that: The results of the analysis provide a deep understanding of the factors that most influence the level of customer satisfaction. The research findings identified and analyzed several key factors influencing customer satisfaction. These results provide valuable insights for companies in directing service improvement efforts and developing more effective strategies. This study also discusses the applicability of the C4.5 algorithm in analyzing customer satisfaction survey data, demonstrating its potential as an effective tool to support data-driven decision-making across various industry sectors. In conclusion, this research makes a significant contribution to understanding and improving customer satisfaction through a sophisticated analytical approach. [17]

Based on research published in the journal "The Topsis Method and the C4.5 Algorithm in Determining Recipients of Direct Cash Assistance," it concluded that: The results of the comparative implementation between the TOPSIS method and the C4.5 Algorithm tested on Rapid Miner software showed that the C4.5 Algorithm performed best with an accuracy of 76.00%, a precision of 82.61%, and a class recall of 90.48% using the Direct Cash Assistance (BLT) data sample. So in this study, this Algorithm Method has a better classification level of method performance compared to the TOPSIS Method. In this case, the C4.5 Algorithm method can be used to calculate the determination of recipients of Direct Cash Assistance (BLT) based on each selected criteria attribute [18].

Based on research in the journal entitled "Comparative Analysis of C4.5 Algorithms and Modified K-Nearest Neighbor (MKNN) for Fungus Classification" concluded that: Based on the results of the study, it can be concluded that the performance comparison between the C4.5 algorithm and Modified K-Nearest Neighbor (MKNN) in classifying fungi shows that the C4.5 algorithm has superior performance. Performance assessment is carried out by comparing the accuracy, precision, recall, and f1-score of the two algorithms. Although both algorithms are able to produce classification models with an accuracy level above 90%, the C4.5 algorithm obtained better results compared to Modified K-Nearest Neighbor (MKNN). In this study, the C4.5 algorithm succeeded in obtaining an accuracy level of 98.52%, precision of 98.55%, recall of 98.52%, and f1-score of 98.51%. In contrast, the Modified K-Nearest Neighbor (MKNN) algorithm using a value of K = 10 achieved an accuracy rate of 96.62%, a precision of 96.69%, a recall of 96.62%, and an f1-score of 96.57%. Of the total of 1625 poisonous and non-poisonous (edible) mushroom data, the C4.5 algorithm successfully classified 1602 data correctly, while the Modified K-Nearest Neighbor (MKNN) algorithm successfully classified 1570 data correctly. The C4.5 algorithm also succeeded in reducing the number of attributes from 23 to 5. This caused the C4.5 algorithm classification model to be simpler in terms of calculation complexity. In contrast, the Modified K-Nearest Neighbor (MKNN) achieved good performance using a value of K = 10. From the results of this study, it can be concluded that the C4.5 algorithm is more effective for classifying mushroom data [19].

Based on research in the journal entitled "Implementation of the C4.5 Algorithm in Coronary Heart Disease Risk Factor Analysis" concluded that: Data pre-processing was carried out by filtering, namely data duplication, no duplicate medical record numbers were found, filtering missing values, cleaning dyslipidemia and family history attributes, and filtering data inconsistencies were not found. The classification process used the C4.5 algorithm with the RapidMiner tool. Then implementing the k-fold cross validation technique by conducting 5 experiments with values of k = 4, k = 5, k = 7, k = 8, and k = 10. Decision tree formation was carried out with pruning-prepruning and without pruning-prepruning. Model performance was higher when pruning-prepruning was activated with the highest performance obtained from a value of k = 8 with an accuracy value of 86.09%, a precision of 82.63%, and a recall of 91.39%. Based on these performance results, the C4.5 algorithm is considered capable of classifying data well [20]. Based on research in the journal entitled "Application of the C4.5 Algorithm in Classifying Daily Consumption Levels of the Community to Reduce Food Waste in MSMEs in Medan City" concluded that: Based on the results of entropy calculations, information gain, and gain ratio, it is known that the production category attribute has the largest gain ratio value so it is chosen as the root in the formation of the decision tree. The resulting decision tree structure shows that the amount of production is the main factor that influences the daily consumption level, while other attributes such as menu type, weather conditions, and operational days play a supporting role in the classification process. Model evaluation using the 10-fold cross validation method shows that the C4.5 algorithm obtained an accuracy value of 57.56%. These results indicate that the model is able to identify patterns of relationships between MSME operational variables and daily consumption levels [21].

Based on research in the journal entitled "C4.5 Algorithm Optimization With Backward Elimination Selection Feature For Creditworthiness Assessment," it was concluded that: Based on the results of research and testing, the performance of the C4.5 model without backward elimination for creditworthiness assessment provided an accuracy level of 91.90% with an area under the curve (AUC) of 0.915. Meanwhile, the performance of the C4.5 model with backward elimination provided an accuracy level of 94.80% with an area under the curve (AUC) of 0.973. This proves that optimization with backward elimination can improve the performance of the classification method used [22].

Based on research in the journal entitled "Comparison of the C4.5 Algorithm and K-NN in Identifying Students with Potential Dropouts," it was concluded that: The results of testing the classification algorithm for the case of predicting student dropouts for the C4.5 algorithm without the addition of the forward selection feature obtained an accuracy of 95.96%, then after the addition of the forward selection feature, the accuracy increased to 96.66% [23].

Based on research in the journal entitled "Application of the C4.5 Algorithm for Classifying Late Insurance Premium Payments," it was concluded that: based on the results of the implementation and testing of the system that had been carried out, a system for classifying late insurance customer premium payments was successfully built. The system can implement the C4.5 algorithm so that it can produce a decision tree and the system can also classify new data according to the classification rules in the decision tree where the classification accuracy level obtained was 88%. The formation of a pattern that classifies customer data using the C4.5 method provides a visualization that can be used as an option in making decisions for company leaders [24].

Based on research in the journal entitled "Selection for Analysis of Key Performance Indicator Achievements Based on Tracer Study (Case Study: Fasilkom Unsika)" concluded that: The conclusion of this study is that the application of the C4.5 algorithm with the feature selection method on the achievement of IKU1 has produced a significant decision model. Feature selection through forward selection and heatmap validation has helped identify critical features such as "Employment Status", "Salary", "Employed Less Than 6 Months", "1.2 x Minimum Wage", and "Type of Workplace" that influence IKU1. In the implementation of the C4.5 decision tree with the KDD method, the best model of the 3 models generated after feature selection is model 5 (scenario B), achieving 98.77% accuracy with high Precision, Recall, and f1-Score metrics. Suggestions for future research include exploring more comprehensive analysis of all IKUs, using methodologies such as CRISP-DM, experimenting with different classification algorithms, and considering other feature selection methods to obtain more optimal results[25].

3. Result and Discussion

A. Data Analysis

The research dataset consists of 500 employee data, including data on both high-performing and low-performing employees. The research criteria include (1) Age (2) Gender (3) Education Level (4) Length of Service (5) Attendance Level (6) Work Discipline (7) Communication Skills (8) Teamwork (9) Job Targets (10) Productivity (11) Supervisor Assessment (12) Job Training. Backward Elimination is a feature selection technique in the classification process that aims to simplify the model by eliminating less influential attributes, thereby increasing efficiency and optimizing model performance.

Table 1. Sample Dataset

No	Masa Kerja	Tingkat Kehadiran	Disiplin Kerja	Kemampuan Komunikasi	Kerja Sama Tim	Target Pekerjaan	Produktivitas	Penilaian Atasan	Pelatihan Kerja	Kinerja
1	3	Tinggi	Baik	Baik	Baik	Tercapai	Tinggi	Baik	Pernah	Baik
2	5	Tinggi	Sangat Baik	Baik	Sangat Baik	Tercapai	Tinggi	Sangat Baik	Pernah	Baik
3	2	Sedang	Cukup	Cukup	Baik	Tidak	Sedang	Cukup	Belum	Kurang
4	7	Tinggi	Baik	Sangat Baik	Baik	Tercapai	Tinggi	Baik	Pernah	Baik
5	1	Rendah	Kurang	Cukup	Kurang	Tidak	Rendah	Kurang	Belum	Kurang
6	10	Tinggi	Sangat Baik	Sangat Baik	Sangat Baik	Tercapai	Tinggi	Sangat Baik	Pernah	Baik
7	4	Sedang	Baik	Baik	Baik	Tercapai	Sedang	Baik	Pernah	Baik
8	6	Rendah	Kurang	Kurang	Cukup	Tidak	Rendah	Kurang	Belum	Kurang
9	3	Tinggi	Baik	Sangat Baik	Baik	Tercapai	Tinggi	Baik	Pernah	Baik
10	5	Sedang	Cukup	Baik	Cukup	Tidak	Sedang	Cukup	Belum	Kurang

Source: (Research Results, 2026)

A. Model Testing
1. Confusion Matrix

Confusion Matrix

accuracy: 89.60%

	true Baik	true Kurang Baik	class precision
pred. Baik	295	27	91.61%
pred. Kurang Baik	25	153	85.96%
class recall	92.19%	85.00%	

Figure 2. Confusion Matrix

Based on the model testing results using the RapidMiner application, shown in Figure 2, the Decision Tree C4.5 algorithm achieved an accuracy of 89.60%. This value indicates that the model is capable of classifying employee performance data with a fairly high level of accuracy.

The confusion matrix results show that 295 employee data points in the "Good" performance category were correctly predicted, while 25 were misclassified as "Poor." Meanwhile, in the "Poor" category, 153 data points were correctly classified, while 27 were misclassified as inappropriate.

In addition to accuracy, model performance can also be measured by precision and recall. The "Good" class achieved a precision of 91.61% and a recall of 92.19%. The "Poor" class achieved a precision of 85.96% and a recall of 85.00%. These results indicate that the model has a fairly good ability to recognize each class in employee performance data. Overall, the test results indicate that the Decision Tree C4.5 algorithm is capable of producing good classification performance in the employee performance prediction process. The resulting model can be used to support the employee performance evaluation process, thus assisting companies in more objective and efficient decision-making.

1. Backward Elimination

PerformanceVector (Performance)

accuracy: 92.80%

	true Baik	true Kurang Baik	class precision
pred. Baik	312	18	94.55%
pred. Kurang Baik	17	163	90.56%
class recall	94.85%	90.06%	

Figure 3. Backward Elimination

Based on the results of model testing using the RapidMiner application after applying the Backward Elimination method, an improvement in the performance of the Decision Tree C4.5 algorithm in predicting employee performance was observed. The accuracy value reached 92.80%, indicating that the model was able to classify data with a higher level of accuracy than before attribute optimization.

The confusion matrix results showed that 312 employee data points in the "Good" performance category were correctly predicted, while 18 were misclassified as "Poor." In the "Poor" category, 163 data points were correctly recognized, while 17 were misclassified as inappropriate.

In addition to accuracy, the model's performance improvement was also evident in the precision and recall values. The "Good" class achieved a precision of 94.55% and a recall of 94.85%. Meanwhile, the "Poor" class achieved a precision of 90.56% and a recall of 90.06%. These values indicate that the model has a fairly good ability to distinguish each employee performance class based on the data used.

The improvement in classification results was influenced by the feature selection process using the Backward Elimination method. This method removes attributes with low contribution to the model, thus optimizing the classification process. Based on the feature selection results, the attributes eliminated included age, gender, and education level. This allowed the model to only use attributes deemed most influential in determining employee performance, such as tenure, attendance rate, work discipline, productivity, and superior assessment.

Overall, the application of the Backward Elimination method improved the performance of the Decision Tree C4.5 algorithm in predicting employee performance. This was evident in the increase in accuracy, precision, recall, and F1-Score after the attribute optimization process.

2. ROC curve

a. C4.5 Decision Tree Algorithm

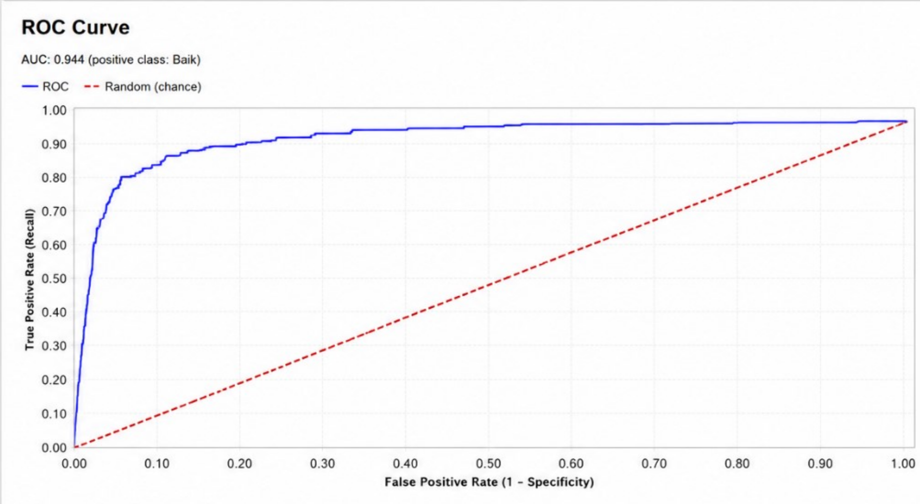

Figure 4. Decision Tree ROC curve

Based on the model testing results using the RapidMiner application, shown in Figure 1, the ROC (Receiver Operating Characteristic) curve indicates that the Decision Tree C4.5 algorithm has excellent classification capabilities in predicting employee performance. This is evident from the AUC (Area Under Curve) value of 0.944, or 94.4%.

An AUC value approaching 1 indicates that the model has a high ability to distinguish employee data categorized as "Good" and "Poor." The ROC curve also shows that the model line is well above the diagonal line (random chance), indicating that the classification results obtained are significantly better than those obtained using random classification.

The ROC graph shows a significant increase in the True Positive Rate (Recall) with a relatively low False Positive Rate. This indicates that the model is able to recognize the majority of high-performing employee data without producing too many classification errors.

These ROC test results support the previous confusion matrix results, which showed an accuracy rate of 89.60%. Thus, it can be concluded that the Decision Tree C4.5 algorithm performs well and is quite stable in predicting employee performance based on the attributes used in this study.

b. Decision Tree C4.5 Algorithm with Backward Elimination

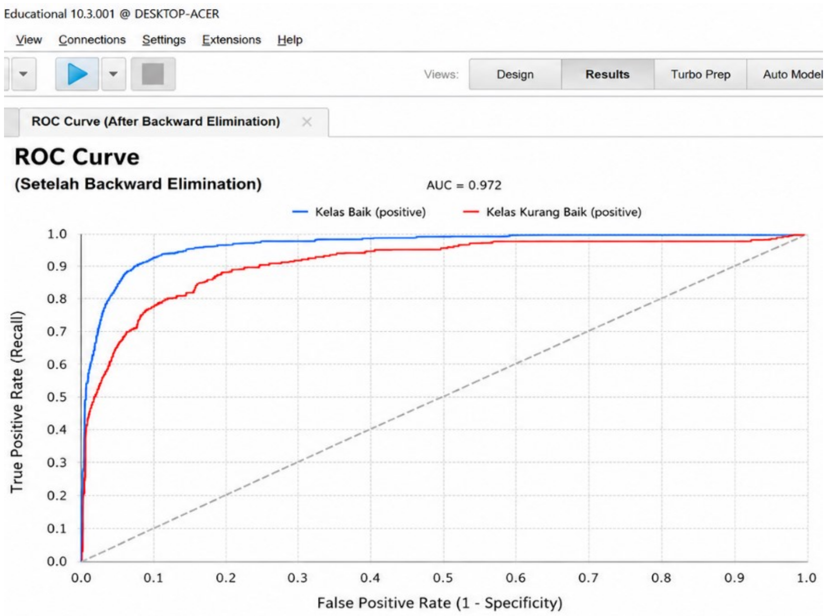

Figure 5. ROC Curve of Decision Tree with Backward Elimination

Based on the results of model testing using the RapidMiner application after applying the Backward Elimination method, the ROC curve in the figure shows an improvement in the performance of the Decision Tree C4.5 algorithm in predicting employee performance. This is evident from the AUC (Area Under Curve) value of 0.972, or 97.2%, indicating the model's classification ability is in the very good category.

The ROC curve shows that the classification line is well above the diagonal line, or random classification. This indicates that the model has a high ability to distinguish employee data in the "Good" and "Poor" performance categories. The closer the ROC curve is to the upper left corner, the better the model's classification performance.

In addition to the AUC value, improvements in model performance are also evident in other evaluation results. The accuracy value achieved was 92.80%, while the precision value for the "Good" class was 94.55%, with a recall of 94.85%. For the "Poor" class, the precision was 90.56% and the recall was 90.06%. These results indicate that the model is capable of recognizing each class with a high degree of accuracy.

The improved model performance was influenced by the feature selection process using the Backward Elimination method. This method successfully eliminated less relevant attributes, thus optimizing the classification process. After attribute elimination, the model used only those attributes that significantly contributed to employee performance prediction.

Overall, the ROC test results indicate that the application of the Backward Elimination method improved the ability of the Decision Tree C4.5 algorithm to distinguish between employee performance data classes. Therefore, the resulting model is considered more effective and can be used to support decision-making in more objective and accurate employee performance evaluations.

3. Results Analysis

Comparison of accuracy, precision, recall, and AUC values for the Decision Tree C4.5 algorithm and Decision Tree C4.5 with Backward Elimination

Table 2. Model Evaluation and Validation

Metode	Accuracy	Precision	Recall	AUC
Decision Tree C4.5	89,60%	91,61%	92,19%	0,944
Decision Tree C4.5 + Backward Elimination	92,80%	94,55%	94,85%	0,972

Source: (Research Results, 2026)

Based on the test results shown in the model evaluation comparison table, the application of the Backward Elimination method improved the performance of the Decision Tree C4.5 algorithm in predicting employee performance. This improvement can be seen in the accuracy, precision, recall, and AUC values obtained after the feature selection process.

The Decision Tree C4.5 model without optimization achieved an accuracy of 89.60%, a precision of 91.61%, a recall of 92.19%, and an AUC of 0.944. These results indicate that the Decision Tree C4.5 algorithm has sufficient classification capability to distinguish between "Good" and "Poor" employee data.

After applying the Backward Elimination method, model performance improved. Accuracy increased to 92.80%, precision to 94.55%, and recall to 94.85%. Furthermore, the AUC value also increased from 0.944 to 0.972. These values indicate that the optimized model has better classification capabilities than the previous model.

Improved model performance is influenced by the process of eliminating attributes less relevant to data classification. By reducing attributes that do not contribute significantly, the model can perform more optimally in recognizing employee performance data patterns. The Backward Elimination method assists in the feature selection process so that only attributes that have a significant impact on the classification results are used in the model.

Overall, the research results show that applying the Backward Elimination method can improve the performance of the Decision Tree C4.5 algorithm in predicting employee performance. This is evidenced by the increase in accuracy, precision, recall, and AUC values after the attribute optimization process.

4. Conclusion

Based on the results of the research that has been carried out, it can be concluded that the Decision Tree C4.5 algorithm can be used to predict employee performance with a fairly good level of classification performance. The resulting classification model can help the employee performance evaluation process more quickly, objectively and efficiently than the manual assessment process.

The test results using the Decision Tree C4.5 algorithm without optimization obtained an accuracy value of 89.60%, precision of 91.61%, recall of 92.19%, and an AUC value of 0.944. This value shows that the Decision Tree C4.5 algorithm has good abilities in classifying employee performance data.

After applying the Backward Elimination method, the model performance increased. The test results show an accuracy value of 92.80%, precision of 94.55%, recall of 94.85%, and AUC value of 0.972. This improvement shows that the Backward Elimination method is able to optimize the performance of the Decision Tree C4.5 algorithm through the feature selection process by removing attributes that are less relevant to the classification process.

Based on the feature selection results, the attributes eliminated include age, gender, and education level. Meanwhile, attributes such as length of service, level of attendance, work discipline, productivity, and superior assessment are the most influential attributes in determining employee performance predictions.

Thus, it can be concluded that the application of the Backward Elimination method has proven effective in improving the performance of the Decision Tree C4.5 algorithm in predicting employee performance. It is hoped that the resulting model can be used to support decision making in the process of evaluating employee performance in companies or agencies.

References

[1] P. W. Sari, G. Giwangga, N. D. Fitriyanti, A. A. Putra, and H. Irawan, "Implementasi Algoritma C4.5 Untuk Klasifikasi Data Insentif Karyawan PT. Adhi Cakra Utama Mulia," *J. Komput. Antart.*, vol. 2, no. 3, pp. 95–101, 2024, doi: 10.70052/jka.v2i3.318.

[2] D. Nasution, A. Marsya, D. Ramatika, R. S. Siburian, and S. Barutu, "Pengujian Algoritma C4.5 Untuk Mengevaluasi Kinerja Pegawai Pada Klinik Lulu," *Surpl. J. Ekon. dan Bisnis*, vol. 2, no. 2, pp. 412–424, 2024, doi: 10.71456/sur.v2i2.879.

[3] M. R. Fanani and E. B. Agustina, "Implementation Of the C4.5 Algorithm With The Backward Elimination Feature Selection For MSME Product Sales Strategy," *J. Artif. Intell. Eng. Appl.*, vol. 3, no. 3, pp. 677–681, 2024, doi: 10.59934/jaiea.v3i3.491.

[4] I. Ubaedi and Y. M. Djaksana, "View of OPTIMASI ALGORITMA C4.5 MENGGUNAKAN METODE FORWARD SELECTION DAN STRATIFIED SAMPLING UNTUK PREDIKSI KELAYAKAN KREDIT.pdf." .

[5] I. Iskandar, L. Hiryanto, and J. Hendryli, "PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA DECISION TREE C4.5 DENGAN TEKNIK PRUNING - PDF Download Gratis.pdf." .

[6] Y. Susanti, M. G. Choyyin, A. Priyatna, and S. Lestari, "Perbandingan Penerapan Algoritma Decision Tree C.45 dan Naïve Bayes Dalam Analisa Kelulusan Siswa pada SMK Swadhipa 2 Natar Kabupaten Lampung Selatan," *J. SIMADA (Sistem Inf. dan Manaj. Basis Data)*, vol. 6, no. 2, pp. 117–123, 2023, doi: 10.30873/simada.v6i2.3772.

[7] M. Mia, A. F. N. Masruriyah, and A. R. Pratama, "The Utilization of Decision Tree Algorithm In Order to Predict Heart Disease," *J. Sisfotek Glob.*, vol. 12, no. 2, p. 138, 2022, doi: 10.38101/sisfotek.v12i2.551.

[8] D. Wilian and S. Sriyanto, "Comparison of the Performance of the C.45 Algorithm with Naive Bayes in Analyzing Book Borrowing at the Library Pringsewu Muhammadiyah University," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 1, pp. 101–106, 2025, doi: 10.32736/sisfokom.v14i1.2300.

[9] A. Y. Yani, F. Ramadhan, D. Irawan, and A. Wasid, "Implementasi Algoritma C4.5 Melalui Pohon Keputusan (Decision Tree) Berbasis Metode Forward Selection Untuk Memprediksi Risiko Kredit Macet Di PT Trust Finance Indonesia Tbk," *REMIK Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 9, no. 4, pp. 1425–1436, 2025, [Online]. Available: https://jurnal.polgan.ac.id/index.php/remik/article/view/15501.

[10] E. Sugiatna, A. M. Ibrahim, and I. Abdul Hadi, "Implementasi Algoritma Klasifikasi C4.5 Untuk Memprediksi Kelayakan Pembelian Kendaraan," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 1, no. 2, pp. 124–132, 2019, doi: 10.35746/jtim.v1i2.26.

[11] M. W. D. Wirawan, S. K. Gusti, J. Jasril, and P. Pizaini, "Penerapan Algoritma C4.5 Mengklarifikasi Penerimaan Bantuan Sosial Menggunakan Feature Selection," *J. Sist. Komput. dan Inform.*, vol. 5, no. 1, p. 69, 2023, doi: 10.30865/json.v5i1.6653.

[12] A. Husnah Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 45–51, 2021.

[13] T. Hidayatulloh, A. Fajria, N. R. Lestari, and N. S. Z. Nufus, "Algoritma C4.5 Untuk Menentukan Kelayakan Pemberian Kredit," *J. Larik*, vol. Vol.2No.2, no. 2, pp. 1–11, 2022.

[14] N. T. Ujianto and N. A. Ramdhan, "Mengukur Tingkat Kepuasan Mahasiswa," vol. 4, no. 01, pp. 105–111, 2022.

[15] D. Yusuf, S. Bahri, and A. Larasati, "Decision Tree Menggunakan Algoritma C4.5 Untuk Analisa Kelayakan Pemberian Kredit," *J. Teknol. Inf.*, vol. 2, no. 2, pp. 97–106, 2021, doi: 10.32546/jutech.v2i2.1660.

[16] I. H. Ikasari and D. Yunita, "Perbandingan Metode Klasifikasi C4. 5 dan Naïve Bayes untuk Mengukur Kepuasan Pelanggan," *J. Inform. Univ. Pamulang*, vol. 6, no. 3, pp. 456–462, 2021.

[17] H. Purba, A. P. Juledi, and I. R. Munthe, "Penentuan Faktor-Faktor yang Mempengaruhi Kepuasan Pelanggan dengan Menggunakan Algoritma C4.5," *J. Ilmu Komput. dan Sist. Informasi(JIKOMSI)*, vol. 7, pp. 91–95, 2024.

[18] N. Ariesanto Ramdhan and H. Bhakti, "Metode Topsis Dan Algoritma C4.5 Dalam Menentukan Penerima

Bantuan Langsung Tunai," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 3, no. 2, pp. 188–196, 2021.

[19] R. Rahmadhani, A. Nazir, F. Syafria, and L. Afriyanti, "Analisis Perbandingan Algoritma C4.5 dan Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Jamur," *J. Sist. Komput. dan Inform.*, vol. 5, no. 2, p. 226, 2023, doi: 10.30865/json.v5i2.7052.

[20] E. A. Yestina, B. H. Prakoso, E. Selviyanti, G. Eko, and J. Suyoso, "Implementasi Algoritma C4 . 5 pada Analisis Faktor Risiko Penyakit Jantung Koroner," *JURIKOM (Jurnal Ris. Komputer)*, vol. 12, no. 5, pp. 684–694, 2025, doi: 10.30865/jurikom.v12i5.8852.

[21] D. Khairani, P. S. Ramadhan, and R. Mahyuni, "Penerapan Algoritma C4 . 5 dalam Klasifikasi Tingkat Konsumsi Harian Masyarakat untuk Mengurangi Food Waste pada UMKM di Kota Medan," vol. 13, no. 2, pp. 569–579, 2026, doi: 10.30865/jurikom.v13i2.9660.

[22] U. Rohwadi, Amrin, and Rudianto, "C4.5 ALGORITHM OPTIMIZATION WITH BACKWARD ELIMINATION SELECTION FEATURE FOR CREDITWORTHINESS ASSESSMENT," *Akrab Juara*, vol. 10, pp. 477–485, 2025.

[23] Y. D. Atma and A. Setyanto, "Perbandingan Algoritma C4.5 dan K-NN dalam Identifikasi Mahasiswa Berpotensi Drop Out," *Metik J.*, vol. 2, no. 2, pp. 31–37, 2018.

[24] J. A. Karlia and W. Nurmansyah, "Penerapan Algoritma C4.5 Untuk Klasifikasi Keterlambatan Pembayaran Premi Asuransi," vol. 09, no. 02, pp. 100–113, 2021.

[25] F. Rahmayani, B. N. Sari, I. Maulana, and R. Mayasari, "Selection Untuk Analisis Capaian Indikator Kinerja Utama Berdasarkan Tracer Study (Studi Kasus : Fasilkom Unsika)," *JATI(Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2732–2738, 2023.