Journal Management, Informatics, Software Engineering and Communication Technology
Vol. 5 No. 1, June (2026) pp. 51-57
Published by: Marq & Cha Institute
Journal Management, Informatics, Software Engineering and Communication Technology
Journal homepage: www.ejournal.marqchainstitute.or.id/index.php/jmirte

Application of Data Mining Using the Random Forest Algorithm to Classify High-Performing Employees at Maju Jaya Computer

Harmoko Lubis[1], Aditiarno Manik[2], Monika Sales Sitompul[3], Sutrisno Situmorang[4]
[1,2,3,4] Manajemen, Akademi Informatika dan Komputer Medicom, Medan, Indonesia

Article Info

Article history

Received : Jun 18, 2026
Revised : Jun 28, 2026
Accepted : Jun 30, 2026

Keywords:

Classification;
Machine Learning;
Outstanding Employee;
Random Forest;
Performance Evaluation.

Abstract

Increasing competition in the business sector requires companies to implement employee performance evaluation systems that are objective, efficient, and accurate. Maju Jaya Komputer, a company engaged in the retail of computer products and related peripherals, requires a performance assessment mechanism capable of processing employee data effectively. The existing conventional evaluation process is time-consuming and prone to subjectivity in identifying outstanding employees. This study aims to implement the Random Forest algorithm to classify outstanding employees based on historical sales records and predefined performance indicators. The research methodology consists of data collection, data preprocessing, Random Forest model development, and performance evaluation using a confusion matrix with accuracy, precision, recall, and F1-score as evaluation metrics. The results demonstrate that the Random Forest model successfully classified employee performance using three key predictor variables, namely the number of laptops sold, attendance rate, and the number of sales prospects. For a new employee record with 26 laptops sold, 94% attendance, and 41 sales prospects, all decision trees consistently predicted the employee as Outstanding, resulting in a majority voting decision of the Random Forest classifier that identified the employee as an outstanding performer. In this prediction example, the model achieved an Accuracy of 100%, Precision of 100%, Recall of 100%, and an F1-score of 100% in classifying outstanding employees. The implementation of this model is expected to enhance the effectiveness of employee performance evaluation while supporting more data-driven human resource management practices.

Corresponding Author:

Harmoko Lubis,
Manajemen Informatika,
Akademi Informatika dan Komputer Medicom,
Jl. Darat No. 74, Kelurahan Petisah Hulu, Kecamatan Medan Baru, Kota Medan, Sumatera Utara 20152, Indonesia.
Email : lubisparholonghian@gmail.com

1. Introduction

The rapid advancement of information technology has intensified competition across various business sectors, including the retail industry for computers and supporting electronic devices. Companies are required not only to offer high-quality products but also to maintain excellent customer service to sustain their competitive advantage. In this context, sales employees play a strategic role as

the primary interface between the company and its customers. Their performance directly influences sales growth, customer satisfaction, and long-term business sustainability (Armstrong & Taylor, 2023).

Maju Jaya Computer, a company specializing in the sales of computer products and related peripherals, relies heavily on the performance of its sales force to achieve organizational objectives. In addition to generating sales revenue, sales employees are responsible for identifying customer needs, providing product recommendations, and building long-term customer relationships. Consequently, the company requires a reliable employee performance evaluation system to monitor productivity, recognize high-performing employees, and support strategic human resource management decisions.

Despite its importance, employee performance evaluation in many organizations is still conducted using conventional approaches that depend heavily on manual assessments and managerial judgment. As business operations expand, the volume of sales transactions, customer prospects, attendance records, and other performance-related data increases substantially. Manual processing of these data is not only time-consuming but also prone to inconsistencies and human bias. Furthermore, traditional evaluation methods often fail to uncover hidden relationships among multiple performance indicators, limiting the effectiveness of identifying outstanding employees (Dessler, 2020).

The increasing availability of organizational data has encouraged the adoption of data-driven decision-making approaches through Artificial Intelligence (AI), particularly Machine Learning (ML). Machine learning techniques have demonstrated significant capability in extracting meaningful patterns from large-scale datasets and producing accurate predictive models for various classification problems. Among numerous classification algorithms, Random Forest has gained considerable attention because of its robustness, high predictive accuracy, and resistance to overfitting. By constructing multiple decision trees and aggregating their predictions through ensemble learning, Random Forest effectively handles heterogeneous datasets while maintaining stable classification performance (Breiman, 2001).

Random Forest has been successfully applied in numerous domains, including employee performance prediction, customer behavior analysis, financial fraud detection, medical diagnosis, and business decision support. Its ability to evaluate multiple performance indicators simultaneously makes it particularly suitable for employee performance classification. In the context of human resource management, the algorithm can analyze variables such as sales achievement, attendance rate, and customer prospect acquisition to objectively identify outstanding employees based on historical performance data rather than subjective managerial assessments.

Based on these considerations, this study proposes the implementation of the Random Forest algorithm to classify outstanding employees at Maju Jaya Computer using historical sales performance data. The proposed approach aims to improve the objectivity, efficiency, and accuracy of employee performance evaluation while providing data-driven recommendations for managerial decision-making. The results of this research are expected to contribute to the development of intelligent decision support systems for human resource management, particularly in employee performance assessment within retail organizations.

2. Research Methodolgy

This study employed a quantitative research approach by implementing the Random Forest algorithm to classify outstanding employees based on historical performance data collected from Maju Jaya Computer. The overall research procedure consisted of data collection, data preprocessing, model development, and performance evaluation.

2.1 Data Collection

Data were collected using two complementary techniques: observation and interviews. Observation was conducted by directly examining employee performance evaluation activities at Maju Jaya Computer. This process enabled the researchers to understand the existing evaluation workflow, identify the performance indicators used by management, and analyze the decision-making process for selecting outstanding employees. The findings from the observation served as the basis for

determining the relevant attributes used in the machine learning model. Interviews were carried out with company stakeholders, including the business owner, sales manager, and human resource personnel. The interviews aimed to obtain detailed information regarding employee performance assessment criteria, the procedures used to determine outstanding employees, and the challenges encountered in the conventional evaluation process. The interview results were subsequently used to validate the selected performance indicators and ensure that the developed classification model reflected actual business requirements.

2.2 Dataset Preparation

The dataset consisted of historical employee performance records obtained from Maju Jaya Computer. Several performance indicators were selected as predictor variables, including the number of laptops sold, attendance rate, and the number of customer prospects generated during the evaluation period. The target variable was employee performance status, which was categorized into two classes: Outstanding and Not Outstanding.

Prior to model development, the dataset underwent preprocessing to improve data quality. This stage included data cleaning, handling missing values, removing duplicate records, verifying data consistency, and transforming categorical attributes into numerical representations when necessary. The cleaned dataset was subsequently divided into training and testing datasets for model construction and evaluation.

2.3 Data Mining Method

This research adopted the Knowledge Discovery in Databases (KDD) framework, which consists of data selection, preprocessing, data transformation, data mining, and knowledge evaluation. During the data mining stage, the Random Forest algorithm was employed as the classification method because of its robustness, high predictive accuracy, and ability to reduce overfitting through ensemble learning.

Random Forest constructs multiple decision trees using bootstrap sampling and randomly selected subsets of predictor variables. Each decision tree independently generates a classification result, and the final prediction is determined through a majority voting mechanism. This ensemble approach improves classification performance and enhances the model's generalization capability for unseen data (Breiman, 2001).

2.4 Random Forest Algorithm

The construction of the Random Forest model follows the standard decision tree learning procedure. For each decision tree, bootstrap samples are generated from the training dataset, while a random subset of predictor variables is selected at every split node. The optimal splitting attribute is determined based on the Information Gain criterion, which is calculated using entropy.

The entropy of a dataset (D) is computed as :

a. Entropy

$$Entropy(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad (1)$$

b. Expected Entropy Setelah Split (Information of Attribute)

$$Entropy_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} Entropy(D_j) \quad (2)$$

3. Information Gain

$$Gain(A) = Entropy(D) - Entropy_A(D) \quad (3)$$

where (Entropy_A(D)) represents the weighted entropy after partitioning the dataset using attribute (A). The attribute with the highest Information Gain is selected as the splitting criterion for each node. The recursive partitioning process continues until a stopping criterion is satisfied, such as reaching the maximum tree depth or obtaining homogeneous class labels. The predictions generated by all decision trees are then aggregated using majority voting to produce the final classification.

2.5 Model Evaluation

The performance of the proposed Random Forest model was evaluated using a confusion matrix. Four commonly used evaluation metrics were employed, namely Accuracy, Precision, Recall, and F1-score. These metrics provide comprehensive measurements of the classification performance by assessing the model's ability to correctly identify outstanding employees while minimizing false classifications. The evaluation metrics are defined as follows:

a. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

b. Precision

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

c. Recall

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

d. F1-Score

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

Description of Variables: D= Dataset, D_j= The j-th data partition, p_i= Probability of instances belonging to class i, m= Number of classes, v= Number of attribute partitions, TP= True Positive, TN= True Negative, FP= False Positive, FN= False Negative.

The resulting evaluation metrics were analyzed to determine the effectiveness of the Random Forest algorithm in classifying outstanding employees and to assess its suitability as a decision support tool for employee performance evaluation at Maju Jaya Computer.

3. Results And Discussions

3.1 Dataset Description

The dataset used in this study consists of historical performance records of ten laptop sales employees at Maju Jaya Computer. Three predictor variables were employed to represent employee performance, namely the number of laptops sold, attendance rate, and the number of customer prospects. The target variable was employee performance status, which was categorized into two classes: Outstanding and Not Outstanding.

Table 1 presents the employee performance

No	Salesperson's Name	Number of Laptops Sold	Attendance (%)	Number of Prospects
1	Julianto	8	80	15
2	Junita	25	96	40
3	Andri Eka Pratama	14	86	24
4	Aris Panjaitan	32	99	50
5	Banjirul Akbar	12	85	20
6	Adisyaputra	30	98	48
7	Dendy Natalius Purba	15	88	22
8	Dermawan Siallagan	28	97	45
9	Dimas Reihan Hervindo	10	82	18
10	Edya Surachman	27	95	43

Table 1 presents the employee performance dataset used throughout this study. Employees with higher sales achievement, better attendance, and a larger number of customer prospects generally belong to the Outstanding class, whereas employees with relatively lower values are categorized as Not Outstanding. This observation indicates that all three variables contribute to distinguishing employee performance levels.

3.2 Random Forest Classification Process

The Random Forest algorithm was implemented by constructing multiple decision trees using randomly selected predictor variables. In this study, three decision trees were generated based on the available performance indicators. The first decision tree utilized the number of laptops sold as the splitting attribute. Employees who sold more than 20 laptops were consistently classified as Outstanding, while employees with sales of 20 units or fewer were classified as Not Outstanding. The entropy of both resulting subsets was zero, producing an Information Gain value of 1.0, indicating a perfect separation between the two classes. The second decision tree employed the attendance rate as the splitting criterion. Employees with attendance rates exceeding 90% were classified as Outstanding, whereas those with attendance rates of 90% or below were categorized as Not Outstanding. Similar to the first tree, this attribute completely separated both classes and produced an Information Gain value of 1.0. The third decision tree used the number of customer prospects as the splitting attribute. Employees with more than 30 customer prospects were classified as Outstanding, while those with 30 or fewer prospects were classified as Not Outstanding. This attribute also resulted in complete class separation, yielding an Information Gain value of 1.0.

The classification results demonstrate that all selected predictor variables possess strong discriminative capability for identifying employee performance within the available dataset.

3.3 Prediction of New Employee Performance

To evaluate the prediction capability of the proposed model, a new employee record containing the following performance indicators was introduced: Number of laptops sold: 26, Attendance rate: 94%, Customer prospects: 41

Each decision tree independently classified the employee as Outstanding, as summarized in Table 2.

Decision Tree	Prediction
Tree 1 (Laptop Sales)	Outstanding
Tree 2 (Attendance)	Outstanding
Tree 3 (Customer Prospects)	Outstanding

Using the majority voting mechanism, the Random Forest classifier produced the final prediction of Outstanding Employee. The consistency among all decision trees indicates that the employee satisfies all performance criteria learned during model training.

3.4 Classification Performance Evaluation

The classification performance of the Random Forest model was evaluated using a confusion matrix. Table 3 presents the classification results obtained from the testing dataset.

	Actual Outstanding	Actual Not Outstanding
Predicted Outstanding	5	0
Predicted Not Outstanding	0	5

Based on the confusion matrix, the evaluation metrics were calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{5 + 5}{10} = 1.00$$

$$Precision = \frac{TP}{TP + FP} = \frac{5}{5} = 1.00$$

$$Recall = \frac{TP}{TP + FN} = \frac{5}{5} = 1.00$$

$$F1\text{-}Score = \frac{2 \times Precision \times Recall}{Precision + Recall} = 1.00$$

The model achieved an Accuracy of 100%, Precision of 100%, Recall of 100%, and an F1-score of 100% on the available testing dataset.

3.5 Discussion

The experimental results demonstrate that the Random Forest algorithm effectively classified employee performance using three performance indicators, namely laptop sales, attendance rate, and customer prospect acquisition. These variables consistently separated the two employee categories, indicating that they are highly relevant for identifying outstanding sales employees.

The ensemble learning mechanism of Random Forest contributes to the robustness of the classification process by combining predictions from multiple decision trees through majority voting. Compared with a single decision tree, Random Forest generally provides higher predictive stability and reduces the risk of overfitting because each tree is constructed from different bootstrap samples and randomly selected predictor variables (Breiman, 2001).

The perfect classification performance obtained in this study is primarily attributed to the characteristics of the dataset, where the predictor variables clearly distinguish outstanding employees from non-outstanding employees. Consequently, all decision trees generated identical predictions, resulting in unanimous majority voting.

Although the obtained evaluation metrics reached 100%, these results should be interpreted cautiously because the dataset contains only ten observations with highly separable characteristics. In practical applications, employee performance data are usually more heterogeneous and contain overlapping attribute values. Therefore, future research should utilize larger datasets collected over longer operational periods and employ cross-validation techniques to obtain more reliable estimates of model generalization performance.

Overall, the findings indicate that Random Forest is a promising approach for supporting objective and data-driven employee performance evaluation at Maju Jaya Computer. The proposed model can assist management in identifying outstanding employees more efficiently while minimizing subjective judgment during the evaluation process.

4. Conclusion

This study successfully implemented the Random Forest algorithm to classify outstanding employees at Maju Jaya Computer based on three performance indicators: the number of laptops sold, attendance rate, and the number of

customer prospects. The experimental results demonstrate that these variables effectively distinguish between outstanding and non-outstanding employees, enabling the model to generate objective and consistent classification results through the majority voting mechanism of multiple decision trees. Using the experimental dataset, the Random Forest model achieved an Accuracy of 100%, Precision of 100%, Recall of 100%, and an F1-score of 100%, indicating that all employee records were correctly classified. These findings suggest that Random Forest is an effective classification method for supporting employee performance evaluation and reducing subjectivity in managerial decision-making through a data-driven approach. Nevertheless, this study has several limitations, particularly the relatively small dataset used for model development and evaluation. Future research should utilize larger datasets collected over longer operational periods, incorporate additional performance indicators, and compare Random Forest with other machine learning algorithms, such as Decision Tree, Support Vector Machine, Naïve Bayes, or Extreme Gradient Boosting (XGBoost), to further evaluate classification performance and improve model generalization.

References

Armstrong, M., & Taylor, S. (2023). Armstrong's handbook of human resource management practice (16th ed.). Kogan Page.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Dua, D., & Graff, C. (2019). UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences. https://archive.ics.uci.edu/ml

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI Magazine, 17(3), 37–54.

Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

Kuhn, M., & Johnson, K. (2019). Feature engineering and selection: A practical approach for predictive models. Chapman and Hall/CRC.

Mitchell, T. M. (1997). Machine learning. McGraw-Hill.

Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.

Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). Introduction to data mining (2nd ed.). Pearson.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann.

Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). SAGE Publications.

Zhang, C., & Ma, Y. (Eds.). (2012). Ensemble machine learning: Methods and applications. Springer. https://doi.org/10.1007/978-1-4419-9326-7