Journal of Engineering, Electrical and Informatics

E-ISSN: 2809-8706
P-ISSN: 2810-0557

Research Article

Application of Data Mining for Talent Performance Analysis Using The C.45 Method In A Case Study of The Human Resource Department of PT. Xyz

Sutisna [1], Tri Wahyudi [2], and Dedi Gunawan [3], Ivan Pradana [4]

1 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (STIKOM CKI) Jakarta; Email: sutisna@stikomcki.ac.id
2 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (STIKOM CKI) Jakarta; Email: triwahyudi@stikomcki.ac.id
3 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (STIKOM CKI) Jakarta; Email: dedigunawan@stikomcki.ac.id
4 Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (STIKOM CKI) Jakarta
* Corresponding Author: sutisna@stikomcki.ac.id

Abstract: Human resource management plays a critical role in supporting organizational performance, particularly in identifying employees with high competency and leadership potential. The process of evaluating employee performance is often conducted manually, which may lead to subjectivity and inconsistencies in decision-making. This study aims to implement the C4.5 decision tree algorithm for talent performance analysis within the Human Resource Department of PT. XYZ. The research utilized employee performance data collected during the 2022–2023 period, including variables such as attendance, achievement, assessment results, service period, and other competency-related indicators. The study adopted the Cross Industry Standard Process for Data Mining (CRISP-DM) framework, consisting of business understanding, data understanding, data preparation, modeling, evaluation, and deployment stages. The C4.5 algorithm was employed to classify employee competencies and generate decision rules based on entropy and information gain calculations. The results indicate that the algorithm successfully identified the most influential attributes affecting employee performance classification, with achievement, assessment, and service period emerging as key determinants. The resulting decision tree provides a systematic and interpretable classification model that supports objective employee evaluation and talent identification. The study demonstrates that the application of data mining techniques can assist organizations in improving the effectiveness of employee performance assessment and human resource decision-making processes.

Keywords: C4.5 Algorithm; Data Mining; Employee Performance; Human Resource Management; Talent Performance.

Received: August 15, 2022
Revised: September 12, 2022
Accepted: September 25, 2022
Published: October 16, 2022
Curr. Ver.: October 27, 2022

1. Introduction

Human resources are one of the most important assets in an organization because they directly influence organizational performance and business sustainability. As companies continue to expand and diversify their business operations, the need for competent employees with strong leadership potential becomes increasingly critical. Identifying employees who possess outstanding competencies and performance is therefore an essential

task for Human Resource Departments (HRD), particularly in organizations seeking to support long-term business growth and competitiveness.

PT. XYZ is a company primarily engaged in the distribution and retail business of electronic devices. In recent years, the company has expanded its business portfolio into several new sectors, including food, beauty, fashion, and healthcare retail. This business expansion requires the company to identify and develop talented employees who can assume leadership roles in newly established business units. Employees with superior competencies and performance, commonly referred to as star talents, play a significant role in determining the success of organizational development and strategic business initiatives.

Performance evaluation is commonly used as a mechanism to assess employee competencies and identify individuals with high potential. Employee performance reflects the knowledge, skills, abilities, and behaviors demonstrated in completing assigned tasks and contributing to organizational objectives. However, evaluating employee performance is often a complex process because multiple factors must be considered simultaneously, including work experience, attendance records, competency development, assessment results, and overall performance evaluations. In many organizations, the evaluation process is still conducted manually, which may introduce subjectivity and inconsistencies in decision-making.

The increasing volume of employee data presents an opportunity to utilize data-driven approaches for improving the objectivity and efficiency of performance assessment. Data mining techniques have been widely adopted to extract meaningful patterns and knowledge from large datasets. Through classification algorithms, organizations can identify hidden relationships among employee attributes and generate predictive models that support decision-making processes. Data mining not only improves analytical capabilities but also helps organizations make more accurate and consistent personnel decisions.

Several previous studies have demonstrated the effectiveness of data mining techniques in human resource management. Research on employee performance prediction, employee retention, talent identification, and workforce analytics has shown that classification algorithms can provide valuable insights for organizational decision-making. In particular, decision tree-based methods have gained considerable attention due to their interpretability and ability to generate understandable decision rules. A previous study on lecturer performance evaluation using the C4.5 algorithm reported that the method achieved a high classification accuracy and successfully identified top-performing lecturers based on multiple assessment criteria. These findings indicate that the C4.5 algorithm is suitable for classification problems involving multiple performance indicators.

Despite the growing application of data mining in human resource management, the process of identifying star talent within PT. XYZ remains largely dependent on manual assessment and subjective judgment. The absence of an objective classification mechanism may affect the fairness and accuracy of employee evaluations. Furthermore, the large number of employees and evaluation criteria increases the complexity of determining the most suitable candidates for leadership development and career advancement.

To address these challenges, this study applies the C4.5 decision tree algorithm to analyze employee talent performance within PT. XYZ. The classification model is developed using employee data consisting of attendance records, competency learning achievements, assessment results, and other relevant performance indicators. The objective of this study is to evaluate the effectiveness and accuracy of the C4.5 algorithm in classifying employee competencies and supporting the identification of high-performing employees. The resulting model is expected to assist the Human Resource Department in making more objective, efficient, and data-driven decisions regarding employee development and talent management.

The contributions of this study include providing a systematic approach for employee competency classification, supporting the identification of high-potential employees, and generating insights that can help organizations improve workforce development strategies. Through the implementation of data mining techniques, the study aims to enhance the quality of human resource decision-making and support sustainable organizational growth.

2. Literature Review

Data Mining

Data mining is a process of extracting useful knowledge, hidden patterns, and meaningful information from large datasets using statistical, machine learning, and artificial intelligence techniques. The primary objective of data mining is to transform raw data into valuable knowledge that can support decision-making processes. Through data mining, organizations can identify relationships among variables, predict future outcomes, and classify data into meaningful categories.

In the context of human resource management, data mining can be utilized to analyze employee performance, identify talent potential, predict employee turnover, and support strategic workforce planning. The increasing availability of employee-related data enables organizations to apply data-driven approaches for improving the effectiveness and objectivity of personnel management decisions.

Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) refers to the overall process of discovering useful knowledge from data. KDD consists of several sequential stages, including data selection, data preprocessing, data transformation, data mining, and interpretation of results. These stages are designed to ensure that raw data can be transformed into meaningful information that supports organizational objectives.

The data selection stage focuses on identifying relevant datasets for analysis. Data preprocessing involves cleaning inconsistent, incomplete, or redundant data. Data transformation converts the prepared data into an appropriate format for mining activities. Subsequently, data mining techniques are applied to discover patterns and relationships within the dataset. Finally, the interpretation stage evaluates and presents the extracted knowledge in a form that can be understood and utilized by decision-makers.

Classification in Data Mining

Classification is one of the most widely used techniques in data mining and belongs to the category of supervised learning. The objective of classification is to assign data instances into predefined classes based on their attributes and historical patterns. A classification model is developed using labeled training data and subsequently used to predict the class of new observations.

Classification techniques have been extensively applied in various domains, including healthcare, education, finance, and human resource management. In employee performance analysis, classification algorithms can be used to categorize employees according to their competency levels, performance achievements, and potential for career advancement.

C4.5 Decision Tree Algorithm

The C4.5 algorithm is a decision tree-based classification method developed as an extension of the ID3 algorithm. It constructs a decision tree by selecting attributes that provide the highest information gain, thereby creating decision rules that can be used for classification purposes.

The C4.5 algorithm evaluates each attribute using entropy and information gain calculations to determine the most influential variable in the classification process. The attribute with the highest gain is selected as the root node, while subsequent branches are formed recursively until all data instances are classified. One of the main advantages of the C4.5 algorithm is its ability to generate interpretable decision rules that are easily understood by decision-makers.

In this study, the C4.5 algorithm is employed to classify employee competencies and identify employees with superior performance based on multiple assessment criteria,

including attendance records, competency learning achievements, assessment results, and other relevant performance indicators.

CRISP-DM Methodology

Cross Industry Standard Process for Data Mining (CRISP-DM) is a widely adopted framework for conducting data mining projects. The methodology provides a systematic approach that guides the entire data mining process from problem identification to model deployment.

CRISP-DM consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Business Understanding focuses on identifying organizational objectives and analytical requirements. Data Understanding involves collecting and exploring available datasets. Data Preparation includes cleaning, transforming, and organizing data for analysis. The Modeling phase applies selected algorithms to develop predictive or classification models. Evaluation assesses the effectiveness of the developed model, while Deployment focuses on implementing the results in practical decision-making environments.

The structured nature of CRISP-DM makes it suitable for human resource analytics projects, particularly those involving employee performance classification and talent identification.

Model Evaluation

The performance of a classification model must be evaluated to determine its effectiveness and reliability. In this study, model evaluation is conducted using a Confusion Matrix and Receiver Operating Characteristic (ROC) analysis.

A Confusion Matrix provides information regarding the number of correctly and incorrectly classified instances. From this matrix, performance metrics such as accuracy, precision, recall, and classification error can be derived. Accuracy measures the proportion of correctly classified instances relative to the total number of observations.

In addition, ROC analysis is used to evaluate the classification capability of the model across different threshold values. The Area Under the Curve (AUC) serves as a quantitative indicator of model performance. An AUC value approaching one indicates excellent classification capability, whereas values closer to zero indicate poor predictive performance. Therefore, Confusion Matrix and ROC-AUC analyses provide comprehensive measurements for assessing the effectiveness of the C4.5 classification model in employee performance evaluation.

3. Materials and Method

Research Design

This study employed a data mining approach to analyze employee talent performance within PT. XYZ. The research aimed to classify employee competencies and identify high-performing employees using the C4.5 decision tree algorithm. The overall research process followed the Cross Industry Standard Process for Data Mining (CRISP-DM) framework, which provides a systematic methodology for transforming organizational data into actionable knowledge.

Data Source and Research Dataset

The dataset used in this study was obtained from PT. XYZ and consisted of employee records collected during the period of 2022–2023. The research focused on active employees who had completed competency assessments, learning modules, certification activities, and internal performance evaluations.

The employee dataset contained several attributes related to employee competency and performance evaluation. These attributes included service period, attendance records, disciplinary violations, competency learning achievements, assessment results, performance evaluations, and other relevant indicators used by the Human Resource Department in evaluating employee performance.

The dataset was utilized as the primary source for constructing a classification model capable of identifying employee competency levels and supporting talent performance analysis.

CRISP-DM Framework

The study adopted the CRISP-DM methodology, which consists of six sequential phases:

Business Understanding

The first phase focused on understanding organizational needs and defining the business problem. PT. XYZ required an objective approach to evaluate employee performance and identify employees with the highest potential for leadership and career development. Therefore, the study aimed to develop a classification model capable of supporting Human Resource decision-making processes.

Data Understanding

At this stage, employee data were collected and examined to understand their characteristics, structure, and suitability for classification purposes. The analysis included identifying relevant variables and evaluating data quality prior to further processing.

Data Preparation

The collected data were prepared before the modeling stage. This process included data cleaning, attribute selection, transformation, and organization of employee records into a format suitable for data mining analysis. Irrelevant or inconsistent data were removed to improve model performance.

Modeling

The modeling phase applied the C4.5 decision tree algorithm to classify employee performance categories. The algorithm generated classification rules based on entropy and information gain calculations. The attribute with the highest information gain was selected as the root node, while subsequent branches were developed recursively until classification rules were established.

The resulting decision tree provided a structured representation of relationships among employee performance indicators and enabled the identification of employees with superior competencies.

Evaluation

The developed classification model was evaluated to determine its predictive capability and effectiveness. Model performance was assessed using classification performance metrics derived from a confusion matrix. The evaluation process measured the ability of the model to correctly classify employee competency categories based on the available dataset.

Deployment

The final phase involved interpreting the classification results and presenting the generated knowledge to support decision-making within the Human Resource Department. The resulting model can be utilized as a decision-support tool for employee competency assessment, talent identification, and workforce development planning.

C4.5 Decision Tree Algorithm

The C4.5 algorithm was selected because of its ability to perform classification tasks using multiple attributes while generating decision rules that are easily interpretable. The algorithm constructs a decision tree by calculating entropy and information gain values for each attribute. Attributes with the highest gain values are prioritized during tree construction because they contribute the most information to the classification process.

Through recursive partitioning, the algorithm produces classification rules that can be used to categorize employees according to their competency and performance characteristics.

Model Evaluation

To evaluate the effectiveness of the classification model, a confusion matrix approach was employed. The confusion matrix provides information regarding correctly and incorrectly classified instances, enabling the calculation of classification accuracy.

In addition, Receiver Operating Characteristic (ROC) analysis and Area Under the Curve (AUC) measurements were utilized to assess overall model performance. A model with an AUC value approaching 1 indicates strong classification capability, whereas lower values indicate weaker predictive performance. These evaluation techniques were used to determine the reliability of the C4.5 model in supporting employee talent performance analysis within PT. XYZ.

Figure 1. Research Framework Based on CRISP-DM Methodology.

4. Results and Discussion

Dataset Characteristics

The dataset used in this study was obtained from PT. XYZ and consisted of employee performance records collected during the 2022–2023 period. The dataset contained employee competency and performance indicators that were utilized by the Human Resource Department to evaluate employee potential and performance. The attributes included service

period, attendance records, disciplinary violations, learning module completion, assessment results, performance evaluations, and other competency-related variables. These attributes were selected because they represent important indicators used in employee performance assessment and talent identification processes.

To support the classification process, the dataset underwent several preparation stages, including data selection, cleaning, transformation, and organization according to the CRISP-DM framework. These processes ensured that the dataset was suitable for classification using the C4.5 decision tree algorithm.

Parameter	Tipe	Keterangan
UniqueId	integer	Tanda nomor sampling
TahunAssessment	integer	Kurun waktu pengambilan
LevelAssessment	integer	Indikator level
Keterangan	string	Penanda tujuan assessment
FormatReport	string	Indikator assessor
ScoreLJ	float	Nilai akhir learning
Learning	string	Status learning
ServicePeriod	string	Masa lama bekerja
Achivement	char	Pencapaian kerja
AchivementFlag	string	Status pencapian kerja
Result	string	Hasil Akhir

Figure 2. Parameter Employee Data mining.

Decision Tree Construction Using the C4.5 Algorithm

The classification model was developed using the C4.5 decision tree algorithm. The algorithm determines the most influential attribute by calculating entropy and information gain values for each variable. The attribute with the highest information gain is selected as the root node of the decision tree.

The initial calculation results indicated that the Achievement attribute produced the highest information gain value of 0.28 and was therefore selected as the root node of the classification model. This finding suggests that employee achievement plays a significant role in determining employee competency classification within PT. XYZ.

After establishing the root node, the algorithm recursively evaluated the remaining attributes to generate subsequent branches. For employees classified under the Achievement = Yes category, further analysis identified Assessment and Service Period as important classification attributes. In particular, the Service Period attribute produced an information gain value of 0.321 in one branch of the decision tree, indicating its substantial contribution to employee competency classification.

Node	Attribute		Jumlah Kasus (S)	Match (S1)	Not Match (S2)	Entropy	Gain
1			319	184	135	0,982913	
	Learning	Total	319	184	135		0,02324904
		Pass	258	160	98	0,957932	
		Not Pass	61	24	37	0,966985	
	Assessment	Total	319	184	135		0,0629443
		Master	24	24	0	0	
		Beginner	295	160	135	0,994813	
	Service Period	Total	319	184	135		0,02933638
		5-6 Years	192	126	66	0,928362	
		3-4 Years	83	36	47	0,987293	
		1-2 Years	44	22	22	1	
	Achievement	Total	319	184	135		0,28946067
		Yes	184	154	30	0,641541	
		No	135	30	105	0,764205	

Figure 3. Initial Entropy and Information Gain Calculation.

Node	Attribute		Jumlah Kasus (S)	Match (S1)	Not Match (S2)	Entropy	Gain
1.1	Achievement Yes		184	154	30	0,641541	
	Learning	Total	184	154	30		0,01042353
		Pass	164	140	24	0,600609	
		Not Pass	20	14	6	0,881291	
	Assessment	Total	184	154	30		0,03451668
		Master	23	23	0	0	
		Beginner	161	131	30	0,693742	
	Service Period	Total	184	154	30		0,03233433
		5-6 Years	126	112	14	0,503258	
		3-4 Years	31	24	7	0,770629	
		1-2 Years	27	18	9	0,918296	

Figure 4. Node Calculation for Achievement.

Node	Attribute		Jumlah Kasus (S)	Match (S1)	Not Match (S2)	Entropy	Gain
1.1.1	Achievement Yes & Assessment Beginner		161	131	30	0,693742151	
	Learning	Total	161	131	30		0,01587749
		Pass	145	121	24	0,647346664	
		Not Pass	16	10	6	0,954434003	
	Service Period	Total	161	131	30		0,3218434
		5-6 Years	110	96	14	0,549910905	
		3-4 Years	27	20	7	0,825626526	
		1-2 Years	24	15	9	0,954434003	

Figure 5. Node Calculation for Achievement = Yes.

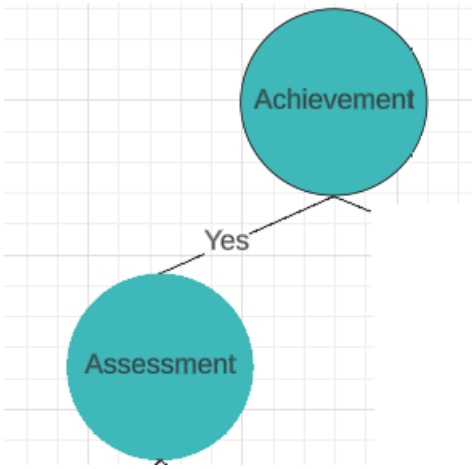

Figure 6. Node Calculation for Achievement = Yes and Assessment = Beginner.

Classification Results

The application of the C4.5 algorithm successfully generated a decision tree model capable of classifying employees according to their competency and performance characteristics. The resulting model transformed complex employee performance data into a set of understandable decision rules that can support Human Resource decision-making processes.

The generated decision tree revealed the hierarchical relationship among employee performance indicators and highlighted the attributes that most strongly influenced classification outcomes. Achievement emerged as the primary classification factor, followed by assessment outcomes and service period. These findings indicate that employee accomplishments, competency evaluations, and work experience collectively contribute to determining employee performance categories.

Node	Attribute		Jumlah Kasus (S)	Match (S1)	Not Match (S2)	Entropy	Gain
1.1.1	Achievement Yes & Assessment Beginner		161	131	30	0,693742151	
	Learning	Total	161	131	30		0,01587749
		Pass	145	121	24	0,647346664	
		Not Pass	16	10	6	0,954434003	
	Service Period	Total	161	131	30		0,3218434
		5-6 Years	110	96	14	0,549910905	
		3-4 Years	27	20	7	0,825626526	
		1-2 Years	24	15	9	0,954434003	

Figure 7. Final Decision Tree Generated by the C4.5 Algorithm.

The resulting classification rules provide practical guidance for Human Resource managers in identifying employees with superior competencies and leadership potential. Furthermore, the model enables organizations to classify employees objectively using historical performance data rather than relying solely on subjective evaluations.

Discussion

The findings demonstrate that the C4.5 algorithm can effectively support talent performance analysis within PT. XYZ. The generated decision tree provides a transparent and interpretable classification mechanism, allowing Human Resource personnel to understand how employee attributes contribute to competency classification.

The identification of Achievement as the root node indicates that employee accomplishments represent the most influential factor in talent classification. This result is consistent with the organizational objective of identifying employees who exhibit strong performance and leadership potential. Additionally, the significance of Assessment and Service Period suggests that competency evaluation and work experience remain important considerations in employee development and promotion decisions.

Compared with conventional manual assessment approaches, the proposed data mining model offers a more systematic and data-driven mechanism for evaluating employee competencies. The classification rules generated by the model can reduce subjectivity in performance assessment and improve consistency in decision-making processes. Consequently, the model can assist the Human Resource Department in identifying high-potential employees, planning competency development programs, and supporting talent management initiatives.

The study also demonstrates the applicability of data mining techniques in human resource analytics. By utilizing historical employee data, organizations can uncover hidden patterns that may not be easily identified through traditional evaluation methods. Therefore, the implementation of the C4.5 algorithm provides valuable support for strategic workforce planning and employee performance management within PT. XYZ.

5. Conclusion

This study implemented the C4.5 decision tree algorithm to support talent performance analysis within the Human Resource Department of PT. XYZ. The research was conducted using employee performance data consisting of several competency-related attributes, including attendance records, achievement indicators, assessment results, service period, and other performance evaluation criteria. The data mining process followed the CRISP-DM methodology, encompassing business understanding, data understanding, data preparation, modeling, evaluation, and deployment stages.

The results demonstrate that the C4.5 algorithm is capable of classifying employee competencies and generating decision rules that can be used to support talent identification and employee performance evaluation. The classification process successfully identified the most influential attributes affecting employee competency assessment, with achievement emerging as the primary factor in the decision tree structure, followed by assessment and service period variables.

The generated decision tree provides an interpretable and systematic approach for analyzing employee performance, enabling Human Resource managers to perform evaluations more objectively and consistently. By utilizing historical employee data, the model assists in identifying employees with superior competencies and supports decision-making related to talent management and workforce development.

Overall, the implementation of the C4.5 algorithm demonstrates its potential as a decision-support tool for employee performance classification and talent identification within PT. XYZ. The findings indicate that data mining techniques can contribute to improving the effectiveness of human resource management by providing a more structured, data-driven, and transparent evaluation process.

References

Al-Radaideh, Q. A., Al Nagi, E., & Baradei, M. (2012). Using data mining techniques to build a classification model for predicting employees' performance. International Journal of Advanced Computer Science and Applications, 3(2), 144–151.

Han, J., Kamber, M., & Pei, J. (2022). Data mining: Concepts and techniques (4th ed.). Morgan Kaufmann.

Hidayah, A. N. Z., & Nugroho, A. (2021). Penerapan data mining dalam menentukan kinerja dosen menggunakan algoritma C4.5. Jurnal Sistem Informasi dan Artificial Intelligence, 1(2), 57–66.

Hijrah, H., Supriadi, S., & Yanti, D. (2022). Comparative analysis of data mining applications in predicting the quality of employee performance using the C4.5 algorithm method. JATISI (Jurnal Teknik Informatika dan Sistem Informasi).

Kusumaningtyas, D. P., & Legowo, N. (2023). Data mining techniques to predict high leadership potential based on talent pool assessment. Journal of System and Management Sciences, 13(6), 336–346.

Liu, J. (2023). Design and application of human resource management system based on data mining technology. Procedia Computer Science, 228, 241–252. https://doi.org/10.1016/j.procs.2023.11.028

Nayem, Z., & Uddin, M. A. (2024). Unbiased employee performance evaluation using machine learning. Journal of Open Innovation: Technology, Market, and Complexity, 10(1), 100243. https://doi.org/10.1016/j.joitmc.2024.100243

Pasaribu, A. B., & Irawan, M. D. (2024). Decision support system implementation of decision tree algorithm C4.5 in employee performance assessment. *Jurnal E-Komtek, 8*(2), 373–386. https://doi.org/10.37339/e-komtek.v8i2.2145

Quinlan, J. R. (1993). C4.5: Programs for machine learning. Morgan Kaufmann.

Saputra, W., & Purwitasari, D. (2022). Fatigue management: Machine learning application for predicting mining worker fatigue. In 2022 International Conference on Information Technology Research and Innovation (ICITRI) (pp. 117–122). IEEE. https://doi.org/10.1109/ICITRI56423.2022.9970203

Suksangaram, W., Hemtong, W., & Klamsakul, S. (2018). Predicting learning organization factors that affect performance by data mining techniques. In *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)* (pp. 1–6). IEEE. https://doi.org/10.1109/iSAI-NLP.2018.8692984

Trisnadewi, N. K. A., & Suputra, G. A. (2023). Pengaruh motivasi kerja, kompetensi dan komitmen organisasi terhadap kinerja pegawai. WidyaAmrita: Jurnal Manajemen, Kewirausahaan dan Pariwisata, 3(6), 1231–1242.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann.