JURNAL BISNIS, MANAJEMEN & PERBANKAN

RESEARCH ARTICLE
Published: September 27, 2023
doi: 10.21070/jbmp.v9i2.1718

Forecasting Employee Potential through Probationary Assessment

Asradiani Novia[1], Imam Yuadi[2]*
1 Department of Human Resources Development, Postgraduate School, Universitas Airlangga, Surabaya, Indonesia
2 Department of Information and Library Science, Universitas Airlangga, Surabaya, Indonesia

Effective corporate governance necessitates the continual nurturing and cultivation of employee potential for long-term professional success. However, assessing an employee's potential and performance objectively and consistently from the start of their career presents a substantial difficulty in reducing any mismatches with the company's goals and expectations. This study introduces a predictive methodology that uses probationary employee performance to map their potential. The study focuses on Performance (Y-axis) and Potential (X-axis) variables using data from 265 employees at Company X who went through a probationary period. Various machine learning models, including Logistic Regression, Naive Bayes, k-NN, SVM, and Decision Tree, were used to analyze data using Orange Data Mining software. The Logistic Regression model has the highest accuracy, at 90% (0.906). Validity testing, using the Confusion Matrix, allowed individuals to be classified into nine potential groups, in accordance with the 9-Box Matrix Talent Management paradigm. This classification provides HR with a strategic tool for tailoring career development strategies based on expected potential within their respective sectors.

Keywords: *Employee Potential, Probationary Period, Predictive Model, Talent Management, Machine Learning Models*

OPEN ACCESS
ISSN 2528-4649 (online)
ISSN 2338-4409 (print)

**Correspondence:*
Asradiani Novia
asradiani.novia-2022@pasca.unair.ac.id

Citation:
Received: May 15, 2023
Accepted: July 25, 2023
Published: September 27, 2023
JBMP: Jurnal Bisnis, Manajemen dan Perbankan.
Vol: 9/ No. 2doi:10.21070/jbmp.v9i2.1718

INTRODUCTION

Human Resources (HR) is essential in business management and achieving organizational goals. Hence employee issues must be addressed to create a conducive work environment (Chung et al., 2023). To have potential and high-performing human resources, collaboration among all parties in an organization is necessary, including managers, team leads, department heads, management, and HR. Employee potential development is also crucial in human resource development (Wahyuningtyas et al., 2021). One of the critical issues in human resource management is the mismatch between the newly recruited employee's potential with the company's needs and expectations, which cannot be identified solely from the recruitment process results (Bright HR, 2019). In Company X, 67% of new employees who have passed the probation period face challenges in adjusting to the work pace, methods, and technologies specific to the company. However, these difficulties are not communicated to managers due to the employees' desire to solve problems independently. As a result, the performance of these employees declines in terms of productivity and motivation.

Apart from that, employee evaluation is only conducted through the annual performance appraisal, determined solely by managers responsible for each employee, without standardized assessment indicators. Managers provide an evaluation in the form of descriptive essays about the employee's job review, submitted via Google Forms, and then processed by the HR department of Company X. This is problematic because it can lead to unfairness in evaluating employee performance. Employees who perceive their evaluations as unfair may experience discomfort at work, resulting in increased terminations or resignations caused by poorly managed knowledge and skills gaps.

The outcome of an employee's probationary period evaluation can serve as a foundation for an early appraisal of an employee's potential, enabling businesses and management to carry out career development and offer the necessary training.(Al-Darraji et al., 2021). Given that Company X already objectively evaluates whether employees are compatible with its expectations during the probation period, the researchers conducted a preliminary study on the HR department of that company to ascertain whether probationary employee evaluations are used as a reference for mapping employee potential. The results showed that out of the 23 HR professionals who responded and worked in the field of HRM, only 21% were aware that probation results could be used to project employee potential. In contrast, 79% believed employee evaluations could only be conducted after performance appraisal. This indicates that 79% of the management and HR need to be made aware of the appropriate analysis and the use of evaluations to map employee potential in the future.

In developing an individual's potential, precise assessments, measurements, and tools are needed, along with better development models than traditional approaches (Garavan et al., 2012). To help manage the best talents in an organization, early identification and mapping of potential are crucial for providing recommendations for future human resource development (career planning). With the early mapping of potential, it is possible to balance the resources needed by the organization and determine future strategies for recruitment, promotion, and both short-term and long-term planning (Susilowati & Fahmie, 2020). Additionally, with technological advancements, HR must innovate by transforming and utilizing analytical data to streamline the analysis of potential employee development, providing a competitive advantage for the company (Baytar, 2022).

Therefore, the early identification of employee potential is essential to maximize their development according to their proper functions. This research provides an overview of utilizing evaluations conducted during the probationary period and after performing the performance appraisal without waiting for a specific period to conduct potential mapping. In this study, the researcher constructed a predictive model using machine learning analysis, where the predictions are generated based on the evaluations conducted during the probationary period. The evaluations provided by the reviewers can offer insights into an employee's potential in the future by observing how the employee performs in each evaluation component during the first three months in a new position at a company. The approaches used are Logistic Regression, Naïve Bayes, k-NN, SVM, and Decision Tree, processed through Orange Data Mining software. This analysis is expected to provide predictive references to HR and management by analyzing the evaluations conducted during the probationary period, aiding in the more objective and practical identification of employee potential.

LITERATURE REVIEW

In this section, the researcher will explain previous research on probationary periods, the gap between the previous research and the one described in this study, and an explanation of the prediction made to be a solution to the previous problem.

Probationary Period

An organization implements a probationary period because the recruitment process does not accurately reflect a candidate's capabilities. According to Matters' Opinion Survey, up to 18% of new hires failed their probationary period, which is determined by how well their performance matched the requirements of their employer over three months (Bright HR, 2019). The probationary period allows HR and company management to assess the effectiveness of hiring new personnel. Identifying employee value can assist in increasing the company's competitive edge. Thus, it will be a better investment if HR understands the worth of the employee themselves rather than competing with rivals (Alziari, 2017).

In this context, HR's problem is to figure out how to recognize an employee's value and potential early. One of the critical responsibilities of human resource management, which is essential to sustaining organizational success, is mapping people's potential regularly (Jain & Nayyar, 2018). Businesses must concentrate more on resolving HR issues and developing productive workplaces (Chung et al., 2023). When executing a strategy to decrease the number of employees that come and depart (employee retention), organizational leaders must have a prediction that can project and map employee potential. It will lower the cost of hiring and training new employees (Qutub et al., 2021).

According to labor regulations in Indonesia based on UU Ketenagakerjaan No. 13 Tahun 2013, the probationary period for employees is an opportunity for the company to assess whether a new employee is suitable for the given job and has the required skills to succeed. This helps the company reduce the risk of recruitment mistakes and ensures that the employed employee will provide a suitable contribution. In addition, the company can evaluate the employee's work abilities and see if they can overcome job challenges and demands (Balfour & Neff, 1993). This evaluation allows the company to determine whether the employee can significantly contribute to the productivity and success of the company.

Furthermore, the probationary period also provides an opportunity for employees to learn more about the company and the job they are doing. They can adapt to the company's culture and values and understand the job's demands and expectations. This can help them to be more successful in the company and contribute more (Dewettinck & van Dijk, 2013). The balance of the workforce is also a consideration for the probationary period because it allows the company to reduce the number of employees who are unsuitable for the company's needs. If an employee does not meet expectations during probation, the company can terminate the employment contract more efficiently and anticipate other risks. Overall, the probationary period is crucial for the company to evaluate the performance of new employees and determine if they are suitable for the position. This helps the company reduce recruitment errors, increase productivity, and ensure that employees are placed in the correct positions (Abedi-Boafo et al., 2019).

Employees Mapping Based on Nine-Box Talent Management

In essence, talent mapping differs from talent management. People management focuses on how each person is expected to perform in line with the business's goals. Everyone has a right to the standards set for output, guidance, and potential development through training. On the other hand, talent is a firm asset in the shape of workers whose requirements are catered to business needs and expected to perform following the company's predictions and strategies (Alziari, 2017).

This potential mapping is usually decided by managers and leaders in companies who hold essential positions to take business strategies to create organizational productivity, where team mapping is based on technical skills (Mabe et al., 2022). The nine-box talent management is a tool used to identify and evaluate employee performance and development potential. Organizations use this tool to map employees into a 3x3 matrix or nine boxes, where the X axis represents current performance and the Y axis represents potential. The 9-Box (Figure 1) concept has been widely adapted and modified by various organizations according to their needs, so various versions and variants exist in its use (McKinsey, 2008). This tool provides a holistic approach to planning the next steps(Academy of Innovate HR & Erik van Vulven, 2020).

This tool plays an important role in determining which talent will be selected by business leaders. It is necessary for performance calibration, employee development planning, determination of benefits or compensation (short-long term compensation), and measuring the company's success or setting targets and OKRs (Objective Key Result). Regarding talent management, two potential matrices must be considered: a. Potential variable on the vertical axis (y) and b. The performance variable is on the horizontal axis (x). These two variable axes categorize employee potential into nine categories. Please see Figure 1.

[Figure 1 about here]

Predictive Model for Mapping Employee's Potential

Today's predictive models have been developed with technological advancements tailored to human job areas. However, predicting employee potential based on probationary period results has not been widely discussed in international journals due to the limitations of probationary period use at the beginning of employment, which is only carried out by a few organizations in countries that implement indefinite employment agreements, including Indonesia. In this case, the researcher wants to analyze the use of employee assessment results during the trial period using Orange Data Mining, which will result in predictive analysis on how to analyze employee potential based on performance during the probation period and group them into a nine-box talent management matrix as a reference for HR and management in developing their potential and talent. The approach used is Logistic Regression, Naïve Bayes, k-NN, SVM, and Decision Tree, which are processed through Orange Data Mining software.

The data analysis produced using this machine learning will make predictions to test the accuracy and precision of the predicted values using these approaches so that HR and business leadership can compare this predictive model with manual potential mapping calculations. In this phase, we will determine the concept and mechanism of each approach used, such as Logistic Regression, Naïve Bayes, k-NN, SVM, and Decision Tree. This approach will compare the accuracy and precision values to guide which predictive model is the most accurate.

Logistic Regression

Logistic Regression is a statistical model used when the dependent variable is binary (only two possible outcomes) or categorical (more than two possible outcomes) (Alduayj & Rajpoot, 2019). A model determines how one or more independent variables—continuous or categorical—will affect the outcome variable or response. The logistic regression model uses the logistic function to map the independent variables to the dependent variable. This function estimates the probability of a specific outcome, considering the values of the independent variables. The logistic regression output is a probability value between 0 and 1 (Jin & Lee, 2017). This probability value is then converted to a binary outcome by setting a threshold, above which the outcome is considered one category and below which the outcome is considered another category.

Naïve Bayes

Naïve Bayes is used to predict the class or label of data based on the features present, assuming that each feature in the data is independent or unrelated. Naïve Bayes is used to predict the probability of a class or label in data based on the occurrence of features in that data. The Naive Bayes method can be used on various data types, such as text, image, and numerical data. One advantage of this method is that it can work quickly and efficiently on huge data sets (Wiratama & Rusli, 2019). Naive Bayes is also easy to implement and can produce good results even with relatively small datasets (Sunarti et al., 2021).

Random Forest

Random forest is an algorithm for classification, regression, and data grouping using an ensemble of decision trees. Decision tree is a data mining model consisting of several decisions or rules that form a tree to predict the output based on the given input (Alabadee & Thanon, 2021). In Random Forest, this algorithm builds many random decision trees and then combines the prediction results from each tree to produce the final prediction. A random subset of the available data and randomly chosen input variables are used in each decision tree. In this process, each tree can produce different results, but prediction errors from each decision tree can be compensated for by combining the prediction results from each tree.

Random Forest has advantages over using only one decision tree. One of them is the advantage of resistance to overfitting or overtraining, which occurs when the model is too complex and too adapts to the training data, making it less general in predicting new data. By using many decision trees, prediction errors that appear on one tree can be compensated for by accurate prediction results on other decision trees.

Support Vector Machine (SVM)

The Support Vector Machine (SVM), a machine learning model, is used not only for multiclass classification but also for classification and regression analysis. The way SVM works is by creating decision boundaries that determine the category of the data and classifying by finding the farthest area boundaries from the data (Vijayarani et al., 2015) for both linear and non-linear data. This model can be used in classifying potential employees as intended by the researcher because it predicts problems when the dependent variable is continuous.

Decision Tree

Decision Tree (DT) is a machine learning algorithm that uses control of continuous attributes and missing values in data. DT is usually represented by a statistical classifier that can be used for grouping nodes (Tu & Chung, 1992). The researcher uses training data to create an outcome tree in this model. The DT algorithm process in machine learning consists of three stages: Construct Model (learning), Evaluation Model (Accuracy), and Model Use (Classification). The classification stage of this research will be based on the percentage of information obtained, in this case, the potential and performance variables evaluated by reviewers when employees are in the probationary period.

METHOD

In this section, we will describe the specific procedures developed and used by the researchers in creating a predictive model according to the objectives of this study. This includes the learning and analysis methodology of the predictive model used, the dataset, variables, data pre-processing, and the model creation process, described in Figure 2.

[Figure 2 about here]

Analysis Methods

In this research, we propose a framework for building and testing a predictive model for mapping employee potential based on assessments during the probationary period. Since data quality is a key factor, researchers need to collect data. The data was collected archive of HR and Management Company X, filled by evaluators who provided performance assessments of employees during the probationary period. The data was tabulated discretely and converted into numeric data through encoding. After preparing the data, it is necessary to shape and select the data type for processing through Orange Data Mining software, where each variable has an ordinal scale.

The next stage is to train the model using the dataset used for prediction. There is a need for data processing, which consists of three subsets: training, validation, and testing. Initially, the researchers clustered employee potential based on probationary results categorized according to the 9-grid/box matrix talent management. This is part of training the model used to validate whether the predictions produced align with the manual mapping. The researchers used a confusion matrix, precision call, F-Score, and AUC for each model to evaluate the data. The best model that resulted in the highest precision would be used as the selected model as the most accurate prediction.

Data Collection and Data Pre-Processing

For data used in this research was a tabulation of evaluations from 265 employees who have undergone a probationary period. The data was collected from archives, and HR documentation of Company X consists of several questions based on its indicator from variable performance and potential.

[Table 1 about here]

Collected data was in a Google form and tabulated data that can be extracted into a .csv file. The raw data had to be cleaned before being processed as part of the data-cleaning process. This data was used according to the research purpose of predicting employee potential based on the evaluation results during the probation period.

The evaluation indicators used consist of discrete data on a Likert scale of 1-4, which will then be transformed into numeric data through encoding.

[Table 2 about here]

In this study, the researchers used machine learning software, Orange Data Mining, to process the data according to the research purpose. Thus, the process of converting employee scores will become a continuous process. The pre-processed data was used for training, testing, and production purposes. 80% of the data was used as training data, and 20% became test data, which was processed using Orange Data Mining. After that, the data was processed in the clustering and model-building phase.

[Table 3 about here]

Model Building

Various approaches can be used to establish clustering based on employee potential and performance. These models will evaluate each strategy's accuracy and precision. Orange Data Mining will be used in this project to create models using logistic regression, random forest, K-NN, decision tree, and SVM techniques to forecast this potential mapping. The data will be grouped using k-Means before the model is constructed. The k-Means method groups arbitrary and random data based on the percentage of nearby data points. The Silhouette score generated by k-Means compares the average distance to elements in other clusters to the distance from the same cluster that is the greatest.

The K-means results are shown by calculating the employee values for each indicator of potential and performance, where this algorithm calculates clusters based on the closest centroid (Ohno, 2022). The stages of this algorithm to find the clusters are to create the closest clusters randomly selected from the training data. The distance calculation is performed using the Euclidean Distance equation as follows:

$$\mathbf{De} = \sqrt{(\mathbf{xi} - \mathbf{si})^2 + (\mathbf{yi} - \mathbf{ti})^2}$$

De is Euclidean Distance, "i" is the number of objects, "x" and "y" are centroid coordinates.

Then this procedure will be repeated several times until convergence is obtained. Calculate the distribution distance (xi, ci) between the training data xi ($1 \leq j \leq M$) and the cluster centroid data cj. The next labeling is done using the Cj method:

$$\mathbf{C_j{:}c_j} = \frac{1}{\mathbf{Cj|}} \sum_i \boldsymbol{\varepsilon\, Cj\, xi}$$

The next step is labeling and adjusting the cluster according to the boxes in the 9 Box Matrix Talent Management model. In essence, the placement of employees according to this 9-box matrix model is determined by the implementation in each company. However, some companies make formulas in calculating employee placement based on the performance appraisal applied. For example, in Company X, the placement of employees according to the matrix box is determined by the following formula:

$$\boldsymbol{X} = (\mathbf{35\%} \times \boldsymbol{P^1}) + (\mathbf{35\%} \times \boldsymbol{P^2}) + (\mathbf{25\%} \times \boldsymbol{P^3})$$

Where :

X = Cumulative Score of the Employee

P^1=sum of Potential Score

P^2=sum of Performance Score

P^3=Sum of culture value Score

Based on Figure 3, the clustering process generated through K-Means will be depicted in a scatter plot that illustrates the distribution of the resulting clusters. At the same time, the results of K-Means will be used to create the prediction model that best approximates accuracy and precision. Therefore, the selected model will also determine the accuracy of implementing the clustering, which is adjusted to the calculation of the 9-Box Matrix of Company X.

[Figure 3 about here]

RESULTS AND DISCUSSION

RESULTS

Implementing machine learning in talent management is one of the ways HR use technologies to simplify their work. Previously, machine learning had also been implemented in the HR field in predicting employee attrition using single machine models such as Decision Trees, Random Forests, and SVM(Jain & Nayyar, 2018). In another case, identify employee potential (Susilowati & Fahmie, 2020) used training results to map employee potential. In addition, mapping the best employee potential (talent mapping) can also be determined with a Social Network Analysis approach (Wahyuningtyas et al., 2021). Those research strengthened the researcher's use of machine learning with a data analytics approach to map employee potential using assessments made during the probationary term as further evidence for this study. Therefore, identifying employee potential can be done early to maximize potential employee development according to the appropriate function.

Labeling

Based on the clustering result from the k-Means, clusters are shown with the silhouette value, which is then depicted through a scatterplot. Then the cluster will be determined and matched with manual calculation from the company to label the cluster. The clustering results show that cluster C5 dominates the division of the 9 clusters. The silhouette value of C5 is around 0.52-0.64. The next largest cluster is Cluster C3, whose silhouette value ranges from 0.44-0.58. If we manually calculate the potential employee mapping based on data from HR of Company X, out of 265 employees, the most proposed is in the Moderate Performance and Potential box with 97 people, and 65 people are in the Moderate Potential and High-Performance box. By comparison, we can infer that the silhouette value of C5 is labeled Moderate Performance and Potential, and the C3 cluster is labeled Moderate Potential and High Performance. The smallest cluster is in Cluster C6, where the number of employees in this category is 12, and they are in the Moderate Performance and Low Potential box. Cluster C1 has a silhouette value of 0.47-0.55 and is labeled as High Potential and Moderate Performance, with 23 employees categorized in this cluster. There are 19 employees in Cluster C2, where the silhouette value ranges from 0.44-0.57, characterized as High Potential and Low Performance. Cluster C4, with six employees, is in the Low Potential and Performance box. Cluster C7 represents

15 employees in the High Performance and Potential box. Cluster C8 is labeled as Moderate Potential and Low Performance due to the number of employees, which is 13 people, and the rest are in Cluster C9, labeled as Low Potential and High Performance.

Labeling on Matrix is presented in Figure 4.

[Figure 4 about here]

Clustering Model

In this analysis, it was found that predicting potential employee analysis using employee assessments during the trial period processed through machine learning showed that Logistic Regression is the model with the highest precision value with a precision of 0.906 or 90.6%, AUC value of 0.995, as well as F1-score, CA, and Recall values of 0.904. Although other models also showed high precision results, such as using Naïve Bayes, in this study, the Logistic Regression model has significantly higher accuracy. Then, to test the precision accuracy of several models used, the researcher added a new file as a testing dataset operated with the generated model. The results from the accuracy test showed that the clustering had shown good accuracy and the data generated as output was correct and precise.

[Table 4 about here]

Based on Table 4, the performance results of each model tested on the training and testing data in this study show that the machine learning model with the highest AUC value is SVM, with a value of 0.994. However, the SVM model has slightly lower precision than the Logistic Regression model, which is 0.895. Meanwhile, the k-NN and Decision Tree models have lower accuracy and precision than other machine learning models.

Choosing the best model to use can be a crucial step in producing accurate and reliable results. To evaluate the performance of a model, the researchers use a confusion matrix to evaluate its performance. A confusion matrix is a table that compares a dataset's predicted and actual values to summarize the categorization outcomes of a model. The Matrix typically contains four values: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). In this research, the confusion matrix shown in Figure 6 shows the Logistic Regression model has the highest accurate and precise model with the lowest accuracy of 83.5% C3 and the highest in C4 with 100% accuracy. Then this model should be good as the predictive model based on cases in Company X.

[Figure 5 about here]

DISCUSSION

Based on the data analytics process above, this predictive model can be implemented as a consideration for an employee following the talent mapping generated from the clustering model and the 9-box matrix talent management principles. With the measurable mapping of employee potential, this methodology becomes a significant problem solver, encompassing general and specific aspects. Mapping potential becomes an urgent matter for companies to define (Kabalina & Osipova, 2022).

This model can be implemented in Company X and various organizations to facilitate management in employee development and performance appraisal according to their potential. This model can be used because it yields high performance (above 96%), as evidenced by research conducted by (Chung et al., 2023). A predictive model can be employed with the best-performing model, resulting in an algorithm that aligns with the expected aspects such as sampling methods, feature selection, data pre-processing, etc. Furthermore, the presence of this predictive model will enhance Corporate Ethical Values (CEV), which assist companies in achieving standards and establishing the boundaries in determining actions and strategies for future human resource management development (Gaudêncio et al., 2014).

1. For employees who are in clusters: C9, C8, C6, C9, and C2, the company should provide maximum learning opportunities that employees can utilize to access specific work-related materials and certifications (Garavan et al., 2012). Such as Udemy Subscription (Professional Account) for all employees and encouragement to pursue professional and skill certifications like AWS Certification.

2. For employees in the clusters of C1, C7, and C3, the company allows managers and supervisors to propose promotions outside of the performance review period if the performance achievement indicators in each quarter exceed expectations and remain consistent. The company should pay more attention to recognizing and retaining employees in these areas as part of relational capital (Torre et al., 2020).

3. For employees in area C4, the company's HR should provide coaching to gather information on how best to develop the potential and performance of these employees from their perspective. Based on previous research, the best recommendation is to consider the employee's values and impact on the organization (Gaudêncio et al., 2014). If improvement is possible, coaching is a good approach. However, if it does not contribute to the company's competitive advantage, disciplinary action may be taken if necessary.

CONCLUSION

In this study, the researchers developed a predictive model using machine learning to help HR and management of Company X to map employee potential early on. The researchers used the evaluation results during the probationary period, which were then analyzed using Orange Data Mining software with several approaches. Employee performance data consisted of two main variables: performance (x) and potential (y), with sub-questions according to the evaluation aspect. The data were then processed to obtain a cluster mapping of employee potential adapted to the 9-box matrix talent management.

After analyzing the data, the Logistic Regression approach model had the highest precision value, 0.906 or 90.6%. To test the accuracy level, the researchers used a confusion matrix and other test performances using recall, F1-score, and AUC values. In building the predictive model, the researchers used the K-Means clustering results to be adjusted with the scatter plot distribution and according to the nine-cluster matrix talent management. The number of employees mapped in the talent management matrix corresponds to Company X's employee potential mapping calculations.

This model can be implemented to help HR analyze employee potential early on. With early potential mapping, the company and management can develop each employee's potential accurately and in line with the employee's interests.

REFERENCES

Abedi-Boafo, E., Duoduaa Nyarko-Tetteh, A., & Tachie-Menson, R. (2019). Assessment of Support Services Available For Staff On Probation in University of Education, Winneba. *In Global Journal of Human Resource Management* (Vol. 7, Issue 2). www.eajournals.org

Academy of Innovate HR, & Erik van Vulven. (2020). The 9 Box Grid: A Practitioner's Guide.

Alabadee, S., & Thanon, K. (2021). Evaluation and Implementation of Malware Classification Using Random Forest Machine Learning Algorithm. *7th International Conference on Contemporary Information Technology and Mathematics, ICCITM* 2021, 112–117. https://doi.org/10.1109/ICCITM53167.2021.9677693

Al-Darraji, S., Honi, D. G., Fallucchi, F., Abdulsada, A. I., Giuliano, R., & Abdulmalik, H. A. (2021). Employee attrition prediction using deep neural networks. *Computers*, 10(11). https://doi.org/10.3390/computers10110141

Alduayj, S. S., & Rajpoot, K. (2019). Predicting Employee Attrition using Machine Learning. *Proceedings of the 2018 13th International Conference on Innovations in Information Technology, IIT* 2018, 93–98. https://doi.org/10.1109/INNOVATIONS.2018.8605976

Alziari, L. (2017). A chief HR officer's perspective on talent management. *Journal of Organizational Effectiveness*, 4(4), 379–383. https://doi.org/10.1108/JOEPP-05-2017-0047

Balfour, D. L., & Neff, D. M. (1993). Predicting and Managing Turnover in Human Service Agencies: A Case Study of an Organization in Crisis. *Public Personnel Management*, 22(3), 473–486. https://doi.org/10.1177/009102609302200310

Baytar, C. U. (2022). Data analytics applied to the human resources industry. In The Future of Data Mining.

Chung, D., Yun, J., Lee, J., & Jeon, Y. (2023). Predictive model of employee attrition based on stacking ensemble learning. Expert Systems with Applications, 215, 119364. https://doi.org/10.1016/j.eswa.2022.119364

Dewettinck, K., & van Dijk, H. (2013). Linking Belgian employee performance management system characteristics with performance management system effectiveness: Exploring the mediating role of fairness. *In International Journal of Human Resource Management* (Vol. 24, Issue 4, pp. 806–825). https://doi.org/10.1080/09585192.2012.700169

Garavan, T. N., Carbery, R., & Rock, A. (2012). Mapping talent development: Definition, scope and architecture. *In European Journal of Training and Development* (Vol. 36, Issue 1, pp. 5–24). https://doi.org/10.1108/03090591211192601

Gaudêncio, P., Coelho, A., & Ribeiro, N. (2014). Organisational CSR practices: Employees' perceptions and impact on individual performance. In International *Journal of Innovation Management* (Vol. 18, Issue 4). World Scientific Publishing Co. Pte Ltd. https://doi.org/10.1142/S136391961450025X

Jain, R., & Nayyar, A. (2018). Predicting employee attrition using xgboost machine learning approach. *Proceedings of the 2018 International Conference on System Modeling and Advancement in Research Trends, SMART* 2018, 113–120. https://doi.org/10.1109/SYSMART.2018.8746940

Jin, S.-B., & Lee, J.-W. (2017). Study on accident prediction models in urban railway casualty accidents using logistic regression analysis model. *Journal of the Korean Society for Railway*, 20(4), 482–490. https://doi.org/10.7782/JKSR.2017.20.4.482

Kabalina, V., & Osipova, A. (2022). Identifying and assessing talent potential for future needs of a company. *Journal of Management Development*, 41(3), 147–162. https://doi.org/10.1108/JMD-11-2021-0319

Mabe, D., Esmael, G., Burg, M., Soares, P., & Halawi, L. (2022). Optimization of Organizational Design. *Journal of Computer Information Systems*, 62(4), 717–729. https://doi.org/10.1080/08874417.2021.1906783

McKinsey, G. (2008, September). Enduring Ideas: The GE–McKinsey nine-box matrix.

Qutub, A., Al-Mehmadi, A., Al-Hssan, M., Aljohani, R., & Alghamdi, H. S. (2021). Prediction of Employee Attrition Using Machine Learning and Ensemble Methods. *International Journal of Machine Learning and Computing*, 11(2), 110–114. https://doi.org/10.18178/ijmlc.2021.11.2.1022

Sunarti, S., Wahyono, I. D., Putranto, H., Saryono, D., Akhmad Bukhori, H., & Widyatmoko, T. (2021). Optimation Parameter and Attribute Naive Bayes in Machine Learning for Performance Assessment in Online Learning. *Proceedings - 4th International Conference on Vocational Education and Electrical Engineering: Strengthening Engagement with Communities through Artificial Intelligence Application in Education, Electrical Engineering and Information Technology, ICVEE* 202. https://doi.org/10.1109/ICVEE54186.2021.9649661

Susilowati, R., & Fahmie, A. (2020). Talent mapping training to improve people analytics efficacy of university students. *Proceedings of the International Conference on E-Learning, ICEL*, 2020-December, 330–334. https://doi.org/10.1109/econf51404.2020.9385483

Torre, C., Tommasetti, A., & Maione, G. (2020). Technology usage, intellectual capital, firm performance and employee satisfaction: the accountants' idea. *TQM Journal*, 33(3), 545–567. https://doi.org/10.1108/TQM-04-2020-0070

Tu, P.-L., & Chung, J.-Y. (1992). A new decision-tree classification algorithm for machine learning. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 1992-Novem, 370–377. https://doi.org/10.1109/TAI.1992.246431

Wahyuningtyas, R., Alamsyah, A., & Diliana, N. A. (2021). Mapping Digital Talent Based on Competency using Social Network Analysis. *Proceeding - 2021 2nd International Conference on ICT for Rural Development, IC-ICTRuDev* 2021. https://doi.org/10.1109/IC-ICTRuDev50538.2021.9656521

Wiratama, G. P., & Rusli, A. (2019). Sentiment analysis of application user feedback in Bahasa Indonesia using multinomial naive bayes. *Proceedings of 2019 5th International Conference on New Media Studies, CONMEDIA* 2019, 223–227. https://doi.org/10.1109/CONMEDIA46929.2019.8981850

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

LIST OF TABLES

TABLE 1 | Variable and Measurement Indicators

Indikator	**Variabel**
Communication Skills	**Potential (Y)**
Accountability	
Initiative	
Tenacy	
Team Work	
Creativity	
Quality of Work	**Performance (X)**
Work Pace	
Independency at Work	
Leadership Skills	
Decision Making	
Work Planning	
Work Delegation	

Source: Data processed, 2023

TABLE 2 | Encoding

Poor	0
Fair	1
Good	2
Excellent	3

Source: Data processed, 2023

TABLE 3 | Pre-Processing and Model Building Settings

Continuize	
Categorical Features	Treat as ordinal
Numeric Features	Leave them as they are
Categorical Outcome(s)	Leave it as it is
K-Means	
Number of Clusters	Fixed : 9
Pre-processing	Normalize Colomn
Re-runs	10
Maximum Iteration	300

Source: Data processed, 2023

TABLE 4 | Model Performance

Model	AUC	CA	F1	Precision	Recall
K-NN	0.951	0.742	0.715	0.753	0.742
Tree	0.846	0.692	0.690	0.697	0.692
SVM	0.994	0.891	0.888	0.895	0.891
Naïve Bayes	0.967	0.777	0.780	0.805	0.777
Logistic Regression	**0.995**	**0.904**	**0.904**	**0.906**	**0.904**

Source: Data processed, 2023

LIST OF FIGURE

FIGURE 1 | Nine-Box Matrix Talent Management

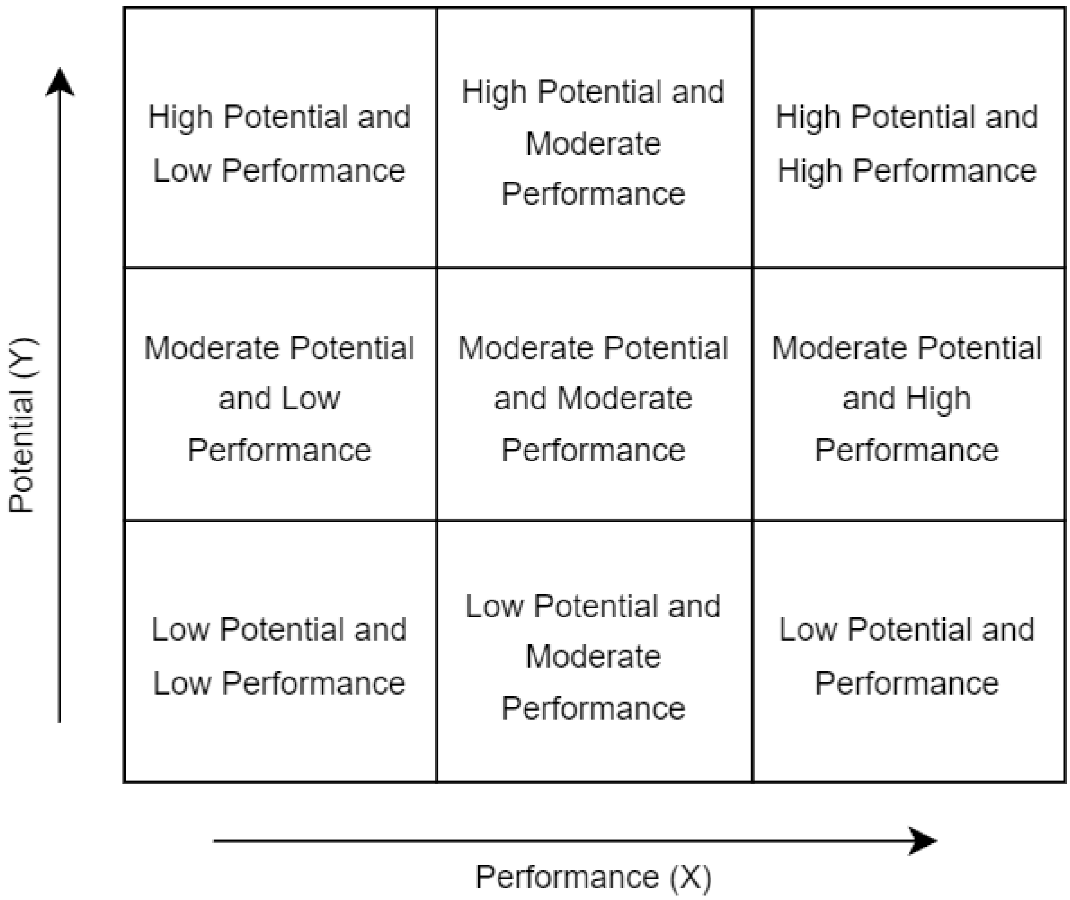

FIGURE 2 | Flowchart of Analytics Process

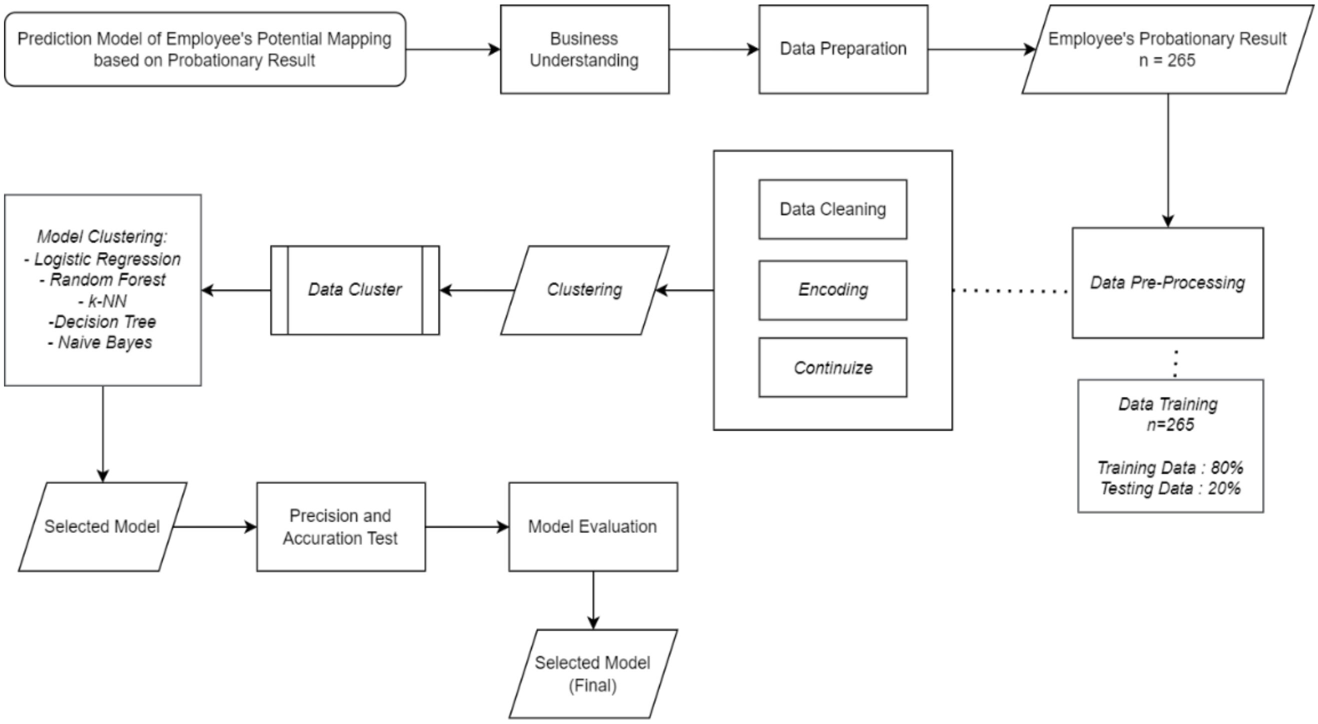

FIGURE 3 | Labeling and Clustering Process

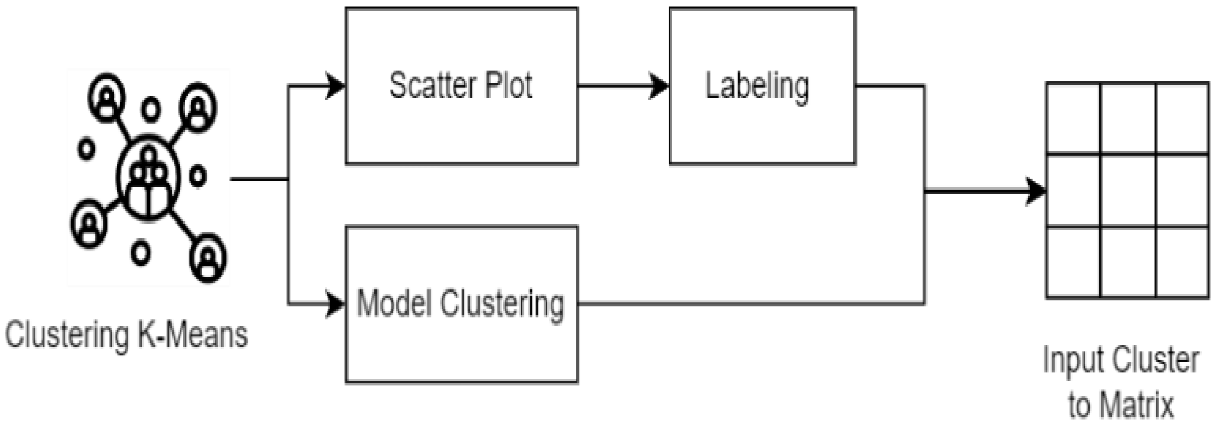

FIGURE 4 | Cluster after Labeling to 9-Box Matrix Talent Management

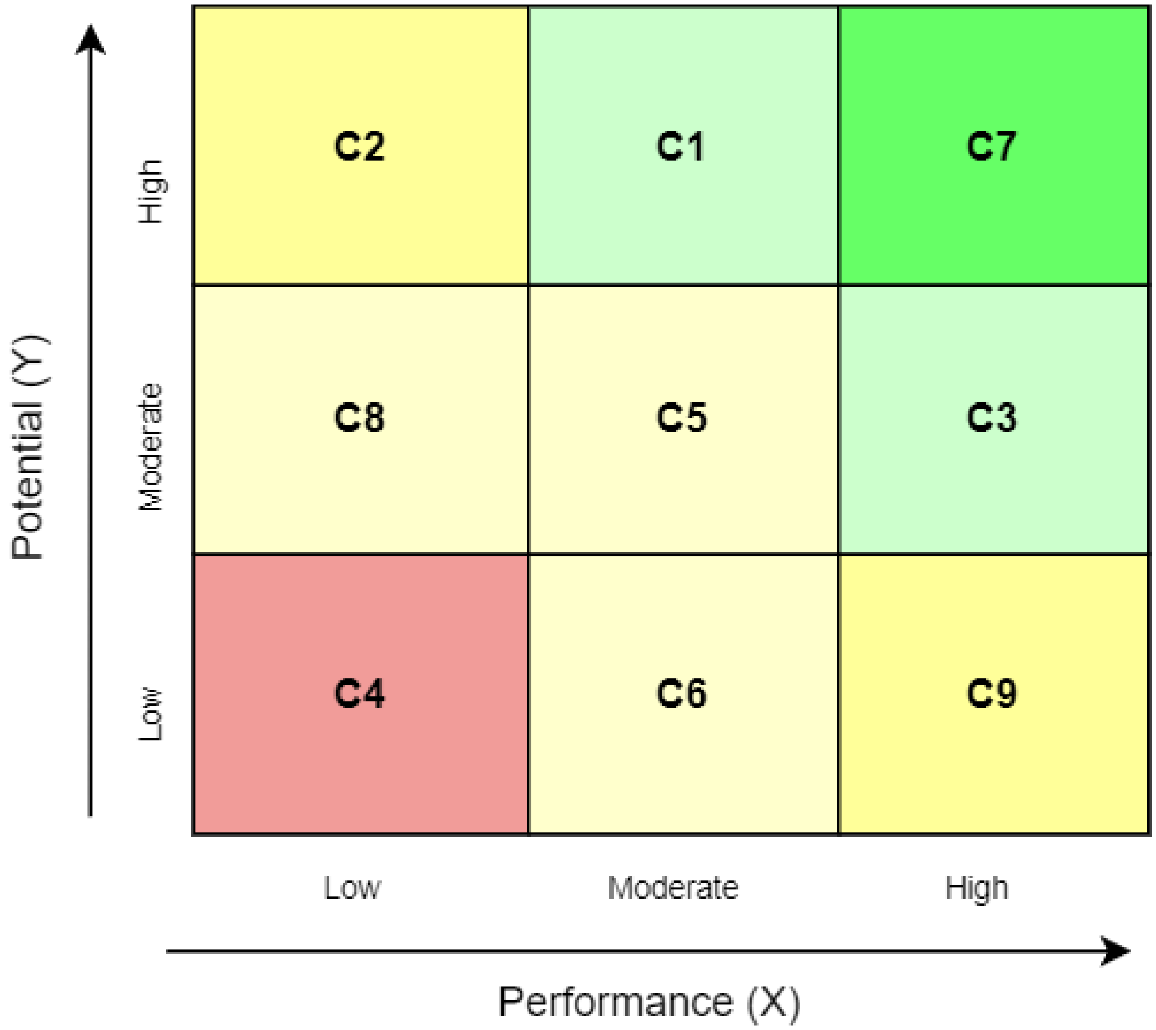

FIGURE 5 | Confusion Matrix of Logistic Regression Model

	C1	C2	C3	C4	C5	C6	C7	C8	C9	∑
C1	89.0 %	0.0 %	4.1 %	0.0 %	1.8 %	0.0 %	0.0 %	0.0 %	0.0 %	80
C2	0.0 %	94.1 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	5.7 %	50
C3	6.1 %	0.0 %	83.5 %	0.0 %	1.8 %	0.0 %	2.2 %	0.0 %	0.0 %	90
C4	0.0 %	0.0 %	0.0 %	100.0 %	0.0 %	0.0 %	8.7 %	8.7 %	2.9 %	10
C5	2.4 %	0.0 %	1.0 %	0.0 %	94.6 %	0.0 %	0.0 %	0.0 %	0.0 %	160
C6	0.0 %	0.0 %	1.0 %	0.0 %	0.0 %	100.0 %	4.3 %	0.0 %	0.0 %	30
C7	2.4 %	0.0 %	6.2 %	0.0 %	1.8 %	0.0 %	84.8 %	0.0 %	0.0 %	50
C8	0.0 %	5.9 %	4.1 %	0.0 %	0.0 %	0.0 %	0.0 %	91.3 %	5.7 %	30
C9	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	85.7 %	30
∑	82	51	97	3	166	27	46	23	35	530