ISSN 2278-3091
Volume 8, No.4, July – August 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse82842019.pdf
https://doi.org/10.30534/ijatcse/2019/82842019

Enhancing Organizational Performance through Employee Training and Development using k-Means Cluster Analysis

Caryl Charlene Escolar-Jimenez, DBA[1], Kichie Matsuzaki, Ph.D.[2], Koji Okada, Ph.D. [3], Reggie C. Gustilo, Ph.D.[4]
[1]Tokyo City University, Japan, caryl@tcu.ac.jp
[2]Tokyo City University, Japan, matsuzak@tcu.ac.jp
[3]Tokyo City University, Japan,okadak@tcu.ac.jp
[4]De La Salle University, Philippines, reggie.gustilo@dlsu.edu.ph

ABSTRACT

K-Means clustering algorithm for machine learning enables organizations to be intuitive by making data-driven decisions and by ensuring effective HR policies and different interventions will add value to an organization. Training and development is an essential component for firms where tangible evidences manifest when employees become more engaged, motivated and committed to contribute to the success of organizations. An assessment of an employee KSAOs needs through k-means clustering to make inferences by grouping similar data to discover underlying patterns to determine the actual deficiencies in the Achievement Category, Leadership Category and Behavior Category to minimize or totally eliminate the human errors that occurs during evaluations. A raw data of 100 employees were evaluated and profiled to determine the proper trainings needed by identifying the skill gaps of individuals. This cluster analysis was able to perform distinct classifications to determine changes in employee training needs to provide valuable insights to both HR and organizational decision makers to understand how employee behavior and needs are changing. This allow firms to be in a better position to respond positively and offer appropriate training that addresses individual staff needs by monitoring employee performance and identifying problem areas in a timely and accurate manner as opposed to the traditional training approaches utilized by many firms. This fast and efficient algorithm allows a straightforward and objective talent inventory that will be the basis for future hires, promotion and retraining decisions.

Keywords: Artificial intelligence, HR analytics, k-means clustering, Management, Machine Learning, Predictive analytics, Training and Development

1. INTRODUCTION

The emergence of HR analytics has tremendously assisted HR departments to quantify and measure organizational success by analyzing data. These vast amounts of data on hand are often underutilized but artificial intelligence platforms have made it easier for HR to systematize and analyze data to identify drivers of business outcomes making them more strategic rather than operational by nature. HR analytics enables HR managers to be more intuitive by making data-driven decisions and by ensuring effective HR policies and different interventions that will further add value to an organization.

Over the years, a number of scholars and practitioners expanded the benchmarking of HR metrics to include investments in training and developing employees [1] that coincides with the comprehensive HR policies and practices of firms. As organizations automate labor intensive tasks, artificial intelligence provides insights to improve the situational awareness that are regularly faced by organizations but these must additionally provide long term actionable business intelligence on how to gain competitive advantage through people.

Training and development is essential in supporting an employee's individual and organizational skills and abilities to increase not only the quality of work provided but these also manage the gaps between an individual's work performance and expectations usually revealed during performance evaluations. Table 1 was referenced from the researchers' previous article wherein a rubric was determined and that each criterion in the performance evaluation will be scored properly by the evaluator by filtering out subjectivity in assessing individual performance [2].

The reliance of organizations with training and development to solve major issues in organizations stem from the lack of knowledge, skills, attitudes and other characteristics (KSAOs) required for tasks that is an integral part of strategy that corporations use to maintain competitiveness [3].

Table 1: Rubric for Assessing Individual Performance

Employee Parameter Rating	Parameter Description
91-100	Outstanding in all categories
81-90	Very Satisfactory except in 1 trait
71-80	Very Satisfactory except in 2 traits
61-70	Very Satisfactory except in 3 traits
51-60	Satisfactory in all traits

41-50	Moderately Satisfactory except in 1 trait
31-40	Moderately Satisfactory except in 2 traits
21-30	Moderately Satisfactory except in 3 traits
11-20	Needs improvement
0-10	Poor

Studies have revealed that as more organizations invest into training, the more effective organizations become in terms of productivity, rate of innovation and rate of turnover. These manifest into tangible evidences when employees become more engaged, motivated and committed to contribute to the success of organizations.

Figure 1 illustrates this training and development article as the third phase in the research series where the evidence approach of HR analytics lean towards the people side of management, utilizes different technologies and methods to reject conventional management approaches with the aim of assisting organizations to perform better by eliminating subjectivity in their decisions.

The objective of this article series is to improve both individual and organizational performance by making better decisions that will contribute to its competitive advantage. These will enhance firm profitability, enlighten management and develop its people by reducing its workload and improving the effectiveness of HR as a whole.

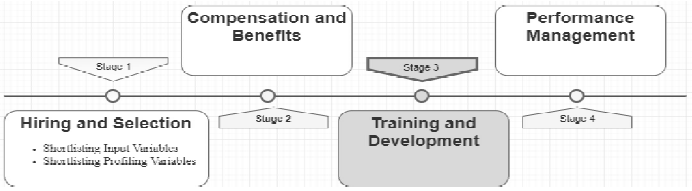

Figure 1: HR Analytics Functional Area Framework

2. PROBLEM STATEMENT

The essence of learning is based on the philosophy that continuous acquisition of learning of knowledge [4] and skills will enable firms to constantly strive for innovation and boost its competitive advantage. Organizations have now devoted more time and resources to training its employees as increased productivity and rate of innovation are just some of the tangible evidences of a more productive and committed workforce. A positive outcome of training is derived from an effective process that starts from an assessment of needs, followed by a careful design and implementation of the training then an evaluation of how well the training achieved the objectives.

However, despite the potential importance of training in organizations, large financial investments spent on training are wasted on poor execution in designing, implementing and evaluating the training and development process. Studies have indicated that usually supervisors or HR allow their subjectiveness to determine their subordinates' training needs with differences in their rating outputs that stem from superficial considerations based on trends or an attractive training brochure. Therefore, it is imperative that organizations must effectively identify the specific training needs to determine where changes are needed and whether the training provided fulfill those intended needs.

The traditional training assessment goes through a tedious and time consuming process by predicting future needs by taking an account all of the current skills and knowledge possessed by current employees through a talent inventory. Then these projections will then identify the employee skill gaps needed for future hiring, promotion and retraining considerations.

A more desirable approach is to base training through careful analysis of employee KSAOs needs through artificial intelligence techniques through k-means machine learning algorithms to make inferences by grouping similar data to discover underlying patterns similar to common applications of clustering in communications [5]. These will determine the actual deficiencies in the Achievement Category, Leadership Category and Behavior Category to minimize or totally eliminate the human errors [6] in decision making.

Furthermore, limited research has been made on the utilization of artificial intelligence platforms needed to correct [7] and addresses employee training that considers the other major HR functional areas specifically, in the hiring and selection and the performance management stages.

With the rapid changes in the technological scenario, global competition, exportation of jobs etc. forces organizations to focus its needs for training and development [8] and at the same time compel organizations to change their strategy to further boost employee performance and productivity rate that will equate to generating more revenue to organizations. Therefore, the crucial role of HR analytics through its evidence-based outcomes will assist firms in achieving their strategic goals by allowing HR departments to make more substantial contributions to the organization as a whole.

.

3. DESIGN PROCESS AND METHODOLOGY

This section provides a detailed discussion on the methodology and process in determining the required level of training needed by each employee based on his or her current performance.

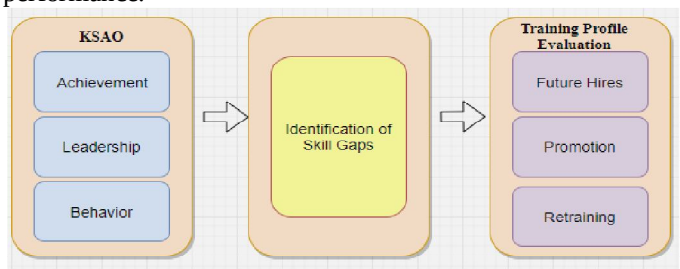

Figure 2: Overview of Employee Training Identification and Evaluation

Figure 2 shows the overall concept of the training identification and profile evaluation that will determine the accurate training requirements of each employee. These will be based on the previous evaluation results acquired from the previous year of tenure. Each work parameter considers the three important categorical profiles namely: achievement, leadership and behavior. The proposed research will utilize K-means cluster analysis to accurately determine the training needs of each employee.

Figure 3 below illustrates the implementation of k-means clustering to determine the level of training that each employee needs to go through.

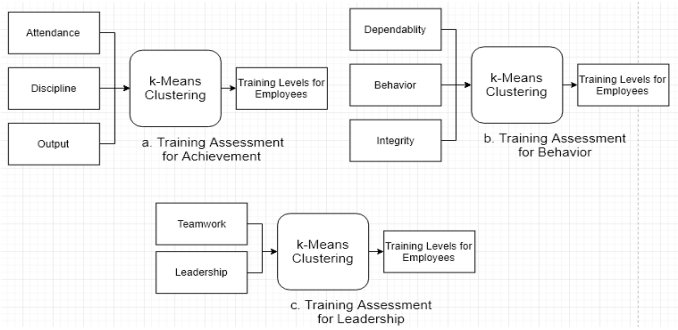

Figure 3: k-means clustering for Achievement, Behavior and Leadership

4. DISCUSSION OF RESULTS

The raw data of 100 employees were evaluated and profiled using the algorithm attributed from the previous employee performance evaluation research. The said preceding evaluation results were used to determine the proper trainings needed by identifying the skill gaps of employees. Each training focuses on three main categories namely: achievement, leadership and behavior. In each categorical area, three levels of competency were proposed in order to fully train and develop the skills of the employees. This will therefore enable HR to prepare a menu of multiple training education corresponding to the employee capability cluster. The levels of competencies on each training area are shown in Table 2 below.

Table 2: Training Levels for Employees

Training Level	Description
1	This introductory training level are for those employees who have incurred poor to very poor evaluations. This level provides basic knowledge and a practical understanding of a range of skill sets.
2	This intermediate training level is designed for those employees exhibiting satisfactory scores in their evaluation. This level is for enhancement of skill sets and provides an in-depth understanding of the organizational environment and related issues.
3	This advanced training level is devised for employees already displaying exceptional performance traits with their work with a previous evaluation score that ranges from very good to excellent. This level explores complex aspects and real-world experiential scenarios of the organizational environment.

The first part of the training plan is the achievement area where the evaluation results of 100 employees reflect the performance traits in the attendance, discipline and output criteria utilizing input parameters through a k-mean clustering. The statistical data was analyzed with a different number of clusters to both identify the homogeneity between the clusters and determine how many training levels are necessary. The elbow curve for the achievement area is illustrated in Figure 4.

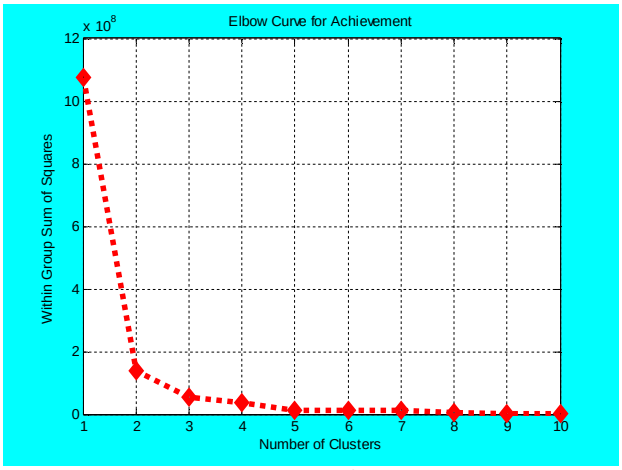

Figure 4: Elbow Curve for Achievement

Figure 4 displayed the elbow curve in the achievement area wherein the elbow point was achieved when the number of clusters is k=2. If 3 clusters were used, there was no significant change in its homogeneity.

To determine the effectiveness of the k-means clustering method, a scatter plot data of the employees is shown in Figure 5 below.

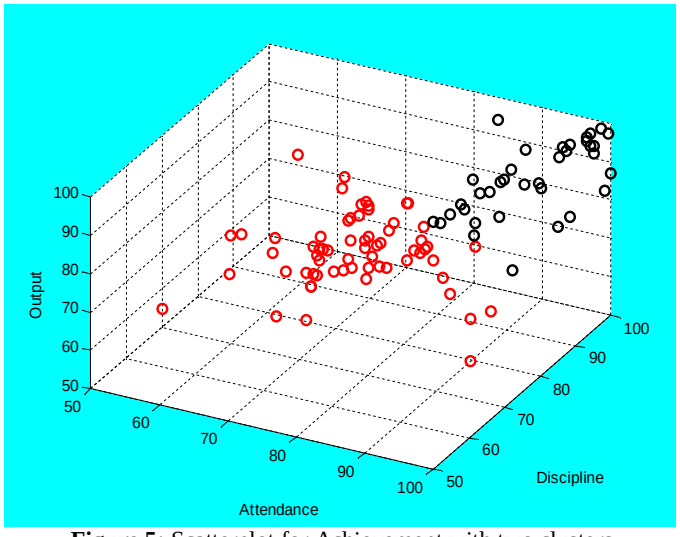

Figure 5: Scatterplot for Achievement with two clusters

Figure 5 presented that the employee performance in the achievement area was clustered into 2 groups. This clearly identifies that the employees were properly grouped according to their evaluation in their attendance, discipline and output traits.

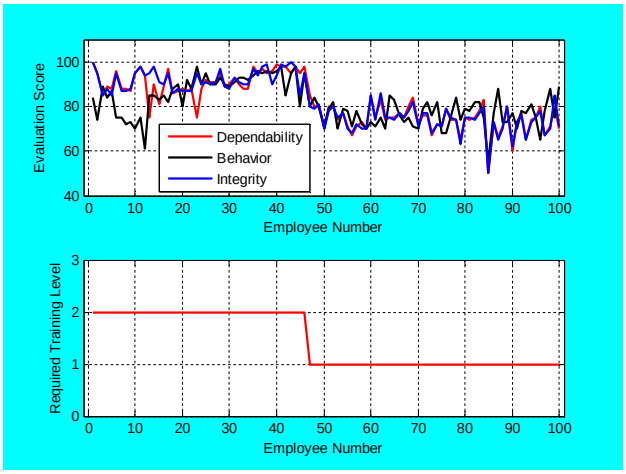

Figure 6: a) Statistical data of 100 employees for their attendance, discipline and output b) Required training level according to their evaluation with two clusters

Moreover, Figure 6 illustrated that 100 employees were properly clustered into 3 levels that determined those employees with high evaluation scores got higher training levels rather than those employees with low or average evaluation scores.

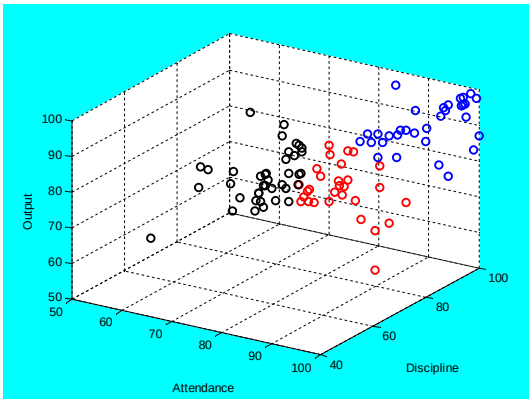

Figure 7: K-means clustering for Achievement area for 100 employees with three clusters

Figure 7 indicated how k-mean was able to divide the statistical data of 100 employees into 3 clusters. Figure 8 below showed the statistical data of the evaluation results for 100 employees according to their attendance, discipline and output traits.

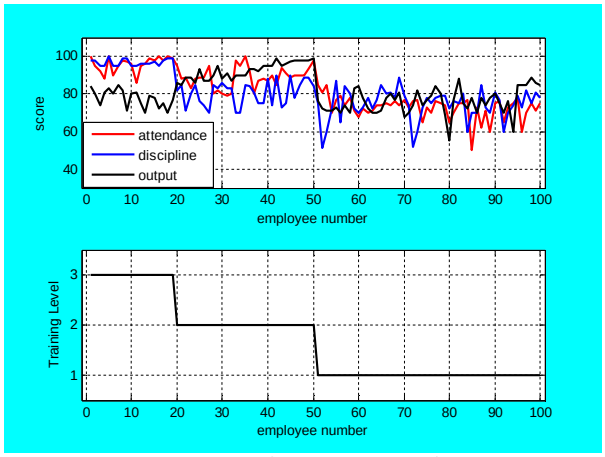

Figure 8: a) Statistical data of 100 employees for their attendance, discipline and output b) Required training level according to their evaluation

Figure 8 showed the consistency of the clustering results in the achievement categorical area from the k-means cluster analysis. Figure 8a presented the evaluation results in the attendance, discipline and output traits of an employee. Figure 8b showed the recommended training level of the employee based on the evaluation results.

The next training plan is for the leadership area. Two important parameters were covered in this category namely: the teamwork ability and the leadership ability of employees. Figure 9 showed the elbow curve of k-mean clustering for the leadership area training plan.

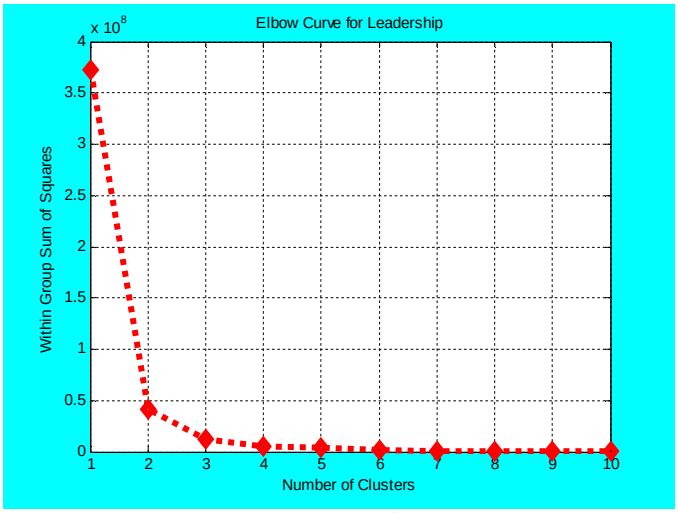

Figure 9: Elbow Curve for Leadership

Figure 9 identified that in the leadership categorical area, the elbow point is achieved when there are 2 clusters. The scatterplots in the statistical data and clustering in the leadership category were presented having 2 and 3 clusters.

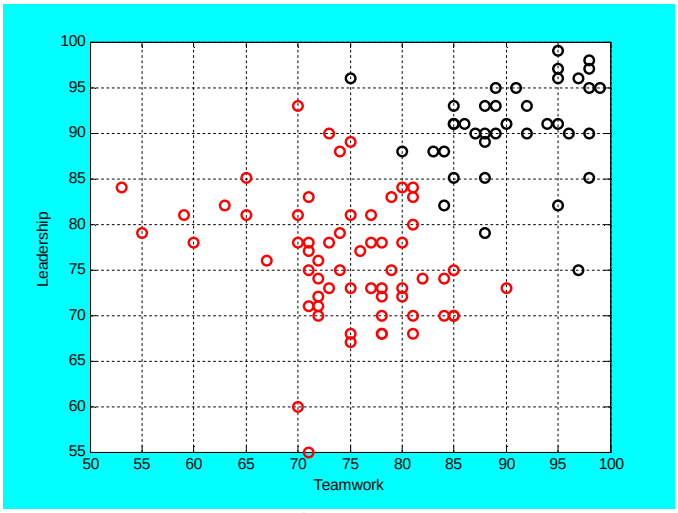

Figure 10: Scatterplot for Leadership with two clusters

Figure 10 and 11 exhibited the grouping of 100 employees in the leadership trait by having 2 clusters. Observe that employees with similar evaluation data are grouped into the same required training level as displayed in Figure 11a and Figure 11b respectively.

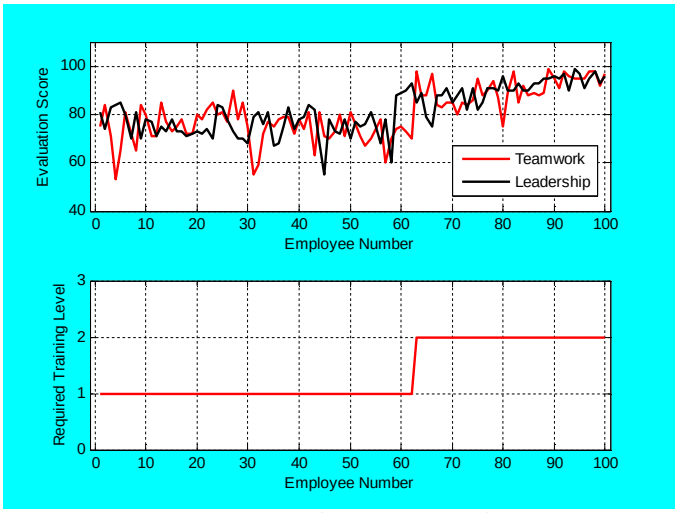

Figure 11: a) Statistical data of 100 employees for their teamwork and leadership b) Required training level according to their evaluation

The clustering of 100 employees in the leadership trait using k=3 or having 3 clusters is presented accordingly in Figure 12 and Figure 13.

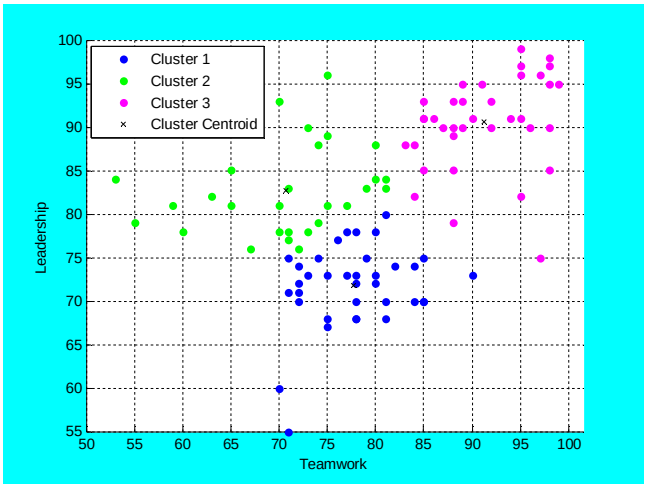

Figure 12: K-means clustering for Leadership area for 100 employees with two clusters

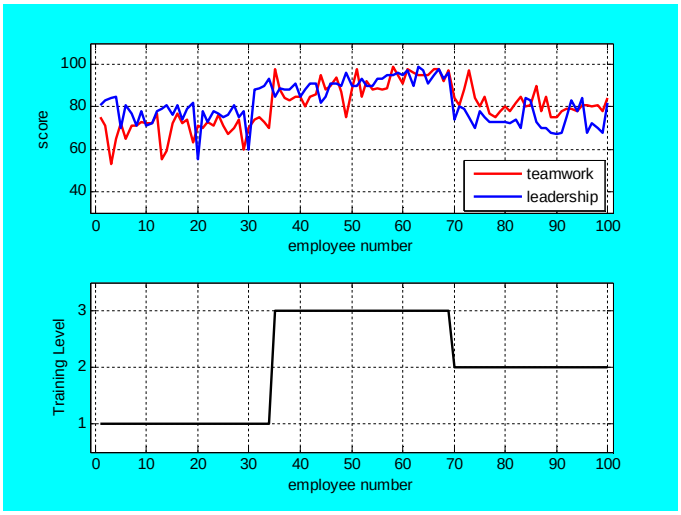

Figure 13: a) Statistical data of 100 employees for their teamwork abilities and leadership abilities b) Required training level according to their evaluation with two clusters

Figure 12 presented the result of dividing the leadership training plan into three levels. The clustering algorithm was forced into 3 clusters for 3 training levels for leadership. The statistical data for this outcome is shown in Figure 13a for the teamwork and leadership traits of the employee. Figure 13b exhibited the recommended training level for each employee.

The third area of the training plan is the behavior area. There are 3 parameters considered in this area namely: dependability, behavior and integrity. These parameters were applied to develop the training plan for the behavior of 100 employees.

The training plan for the behavior categorical area was forced to achieve 3 training levels. The optimal result of the k-means clustering algorithm was revealed in Figure 14 below. The elbow point for the behavior trait is also achieved when k=2 or when there are two clusters.

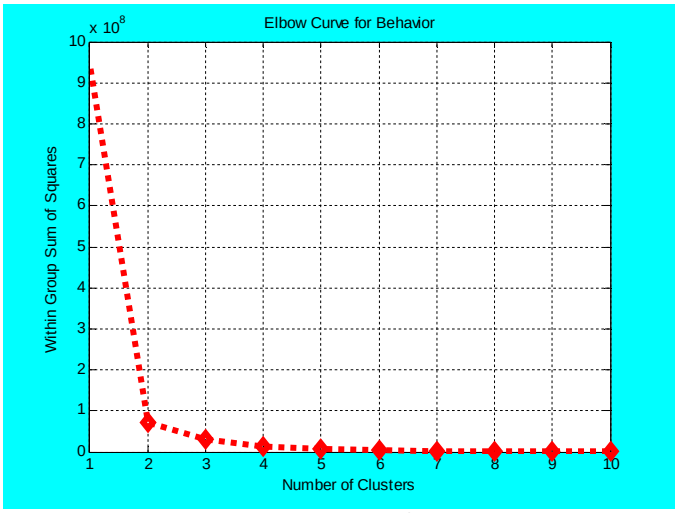

Figure 14: Elbow Curve for Behavior

The clustering of 100 employees for the behavior trait using 2 clusters is shown in Figure 15 below.

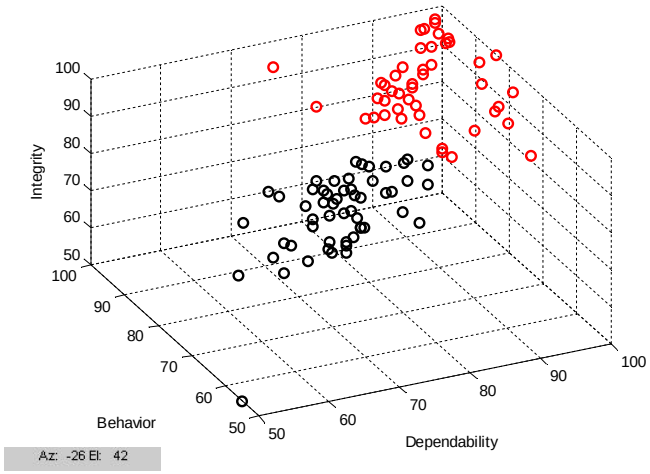

Figure 15: Scatterplot for Behavior with two clusters

Figure 15 and 16 revealed that the grouping of 100 employees for the behavior trait having 2 clusters. Notice that employees with similar evaluation data are grouped accordingly into the same required training level as identified in Figure 16a and Figure 16b respectively.

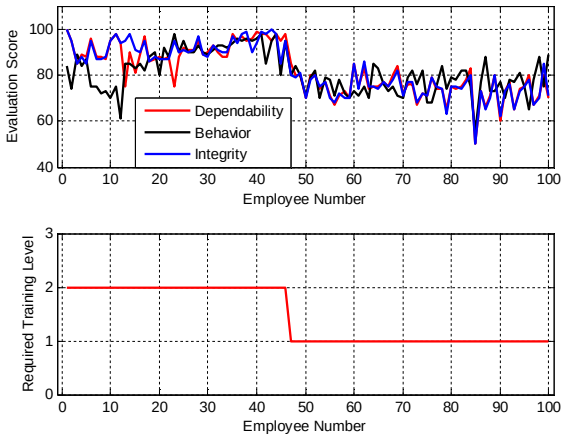

Figure 16: a) Statistical data of 100 employees for their teamwork and leadership b) Required training level according to their evaluation

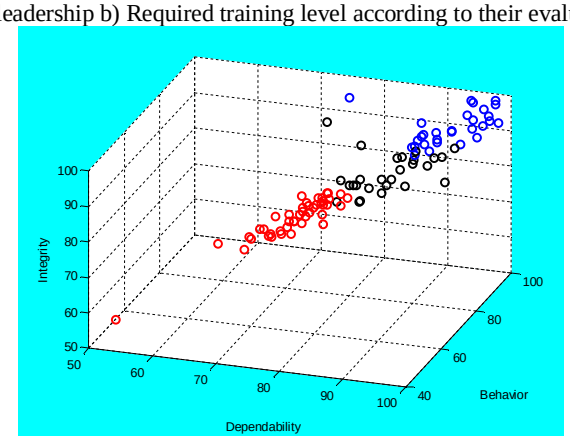

Figure 17: K-means clustering for behavior area for 100 employees with 3 clusters

Figure 17 demonstrated the best result for k-means clustering in the behavior categorical area. The said result minimizes the misplaced clustering as illustrated on the negative part of the silhouette values of the cluster. Additionally, the statistical data for the behavior area is shown in Figure 18a for 3 employee traits namely dependability, behavior and integrity. Figure 18b illustrates the corresponding training level recommended for each employee.

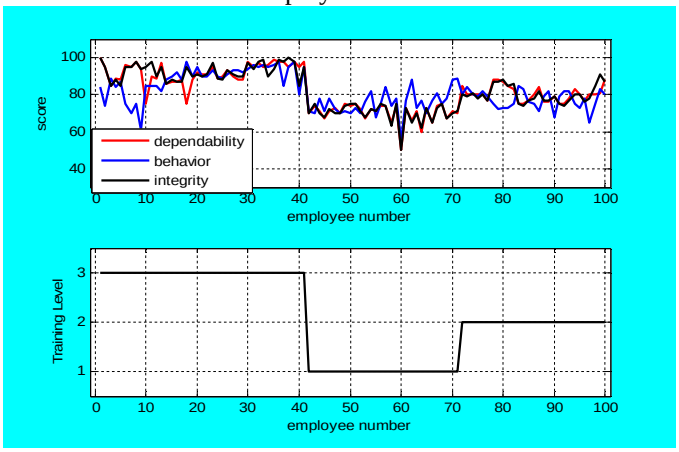

Figure 18: a) Statistical data of 100 employees for their dependability, behavior and integrity b) Required training level according to their evaluation

Overall, the performance of this algorithm [9] indicated that the current talent inventory is a balanced mix of individuals comprised of various skill set levels. With this consideration, HR would be able to truly identify the training education categories that requires 2 or 3 training levels from the cluster analysis.

This educational training system provides a cost benefit to HR to enable them to both effectively prepare a large number of educational training courses when there is a difference in the level of training needs of individuals and becomes more cost effective for them to consolidate trainings into a small number of educational courses when there is no significant difference in the employee level. Therefore, the achievement of this study is that this cluster analysis can reasonably determine the appropriate number of training levels needed by employees.

5. CONCLUSION

The results of the employee performance evaluation served as the basis to construct a comprehensive [10] assessment for the improvement and skill development of employees. Three broad categorical areas were considered for training namely: achievement area training, leadership area training and the behavior area training. The k-means clustering algorithm was utilized to optimally cluster the employees and recognize patterns [11] according to their specific training needs identified as their skill gaps.

This cluster analysis was able to perform distinct classifications to determine changes in employee training needs in a certain time because if data points move from one centroid to another in a given period, this will then provide valuable insights to HR and organizational decision makers to understand how employee behavior and needs are changing.

The results of the k-means clustering is best used for large data and the data results revealed consistent scores that demonstrated the exact training needs of employees. This allow firms to be in a better position to respond positively and offer appropriate training that addresses individual staff needs by monitoring employee performance [12] and identifying problem areas in a timely and accurate manner as opposed to the traditional training approaches utilized by many firms.

Therefore, this fast and efficient algorithm allows a straightforward [13] and objective talent inventory that will be the basis for future hires, promotion and retraining decisions of organizations to further motivate performing employees who are deserving of a promotion or reassignment and provide HR realistic expectations to continuously develop employee skill sets.

REFERENCES

[1] Bassi, L. **Raging debates in HR analytics**. *People and Strategy*, Vol 34.2, pp. 14-11, 2011.

[2] Escolar-Jimenez, C., Matsuzaki, K., & Gustilo, R. **A Neural-Fuzzy Network Approach to Employee Performance Evaluation**. *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 8.3, pp. 573-581, 2019. https://doi.org/10.30534/ijatcse/2019/37832019

[3] Dipboye, Robert L. **Employee Training and Development**. *The Emerald Review of Industrial and Organizational Psychology*, pp. 581-624, 2018. https://doi.org/10.1108/978-1-78743-785-220181016

[4] P. Hanpinitsak, K. Saito, J.-I. Takada, M. Kim and L. Materum, **Multipath clustering and cluster tracking for geometry-based stochastic channel modeling**, IEEE Trans. Antennas Propag., vol. 65, pp. 6015-6028, 2017. https://doi.org/10.1109/TAP.2017.2754417

[5] Chunn,A., & Naugarhiya, A. **Use of Open Source CAD Tools in VLSI Design Curriculum for Developing Countries**. *International Journal of Emerging Trends in Engineering Research*, Vol. 6.4, pp. 15-18, 2018.

[6] L. Villanueva and R. C. Gustilo. **Artificial neural network based antenna sensitivity assignments for chaotic Internet Service Provider network architecture**, Int. J. Eng. Technol. (UAE), vol. 7, pp. 14-17, 2018. https://doi.org/10.14419/ijet.v7i2.3.9958

[7] R. C. Gustilo and E. P. Dadios. **Machine vision support system for monitoring water quality in a small scale tiger prawn aquaculture**, J. Adv. Computat. Intell. Intell. Informat., vol. 20, pp. 111-116, 2016. https://doi.org/10.20965/jaciii.2016.p0111

[8] Momin, W.Y.M. and Mishra, K. **HR analytics as a strategic workforce planning**. International Journal of Applied Research, Vol. 1.4, pp.258-260, 2015.

[9] P. Hanpinitsak, K. Saito, J. Takada, M. Kim and L. Materum, **Clustering method based on scatterer locations for indoor dynamic MIMO channel**, 10th European Conf. Antennas Propag., EuCAP, 2016. https://doi.org/10.1109/EuCAP.2016.7481586

[10] D. D. N. Abinoja and L. Y. Materum, BIC-based optimization of the identification of multipath propagation clusters in MIMO wireless systems, ISAP 2016 - Int. Symp. Antennas Propag., pp. 428-429, 2017.

[11] M. K. Cabatuan, E. P. Dadios and R. N. G. Naguib, **Computer vision-based breast self-examination palpation pressure level classification using artificial neural networks and wavelet transforms**, Proc. IEEE Reg. 10 Annu. Int. Conf. (TENCON), 2012. https://doi.org/10.1109/TENCON.2012.6412282

[12] Pujari, J.D., Bhadangkar, D.K., & Yakkundimath, R. **Identification and Recognition of Facial Expressions Using Image Processing Techniques: A Survey**. *International Journal of Emerging Trends in Engineering Research*, Vol. 5.5, pp. 1-10, 2017.

[13] Africa, A. D. M. **A rough set-based expert system for diagnosing information system communication networks.** International Journal of Information and Communication Technology, Vol. 11.4, pp. 496-512, 2017. https://doi.org/10.1504/IJICT.2017.10008315