Analisis Faktor yang Berkontribusi Terhadap Pengurangan Karyawan Berdasarkan *Clustering Self-Organizing Map*

http://dx.doi.org/10.28932/jutisi.v11i2.11224

Riwayat Artikel

Received: 09 Februari 2025 | *Final Revision*: 23 Juli 2025 | *Accepted*: 23 Juli 2025

Rony Arifiandy✉ [#1], Wiranto Herry Utomo[*2]

[#1] *S2IT - Informatics Study Program, President University*

Jl. Ki Hajar Dewantara, Kota Jababeka, Bekasi 17550, Indonesia

[1]*rony.arifiandy@president.ac.id*

[2] *wiranto.herry@president.ac.id*

✉Corresponding author: ronyarifiandy@gmail.com

Abstrak — **Pergantian karyawan dapat mengganggu operasional organisasi dan sedikit banyak menimbulkan kerugian bagi bisnis. Oleh karena itu, penting untuk memahami faktor-faktor penyebab agar organisasi dapat mengambil tindakan antisipatif. Mengidentifikasi alasan karyawan meninggalkan pekerjaannya sangat penting bagi pemberi kerja dan pembuat kebijakan, terutama jika tujuannya adalah untuk mencegah hal ini terjadi. Data penyebab *turnover* karyawan merupakan data kompleks yang dapat mempunyai banyak dimensi sehingga diperlukan suatu metode tertentu untuk menganalisisnya. Dalam penelitian ini akan dilakukan analisis data penyebab *turnover* karyawan dengan 10 dimensi dengan menggunakan metode *Self Organizing Map* (SOM). *Self-Organizing Map* (SOM) adalah teknik untuk mengelompokkan dan memvisualisasikan data berdimensi tinggi dengan memetakannya ke ruang dua dimensi dengan tetap menjaga struktur topologi data. Metode berbasis jaringan saraf ini memastikan bahwa titik data serupa tetap berdekatan dalam representasi 2D yang dihasilkan. *Self-Organizing Map* akan mengelompokkan data menjadi beberapa kelompok yang seragam. Hasil pengelompokan *Self-Organizing Map* ini akan dinilai dengan skor *Silhouette*, indeks Dunn dan nilai Konektivitas untuk mengetahui seberapa seragam pengelompokannya. Diharapkan dengan menggunakan hasil pengelompokan *Self-Organizing Map* ini menunjukkan bahwa *cluster* yang terbentuk sangat bagus dan datanya terkelompok dengan jelas. Oleh karena itu, kami dapat menganalisis kelompok-kelompok ini dengan hasil yang lebih akurat.**

Kata kunci— ***Clustering***; ***Data Analysis***; ***Data Mining***; ***Self-Organizing Map***; ***Silhouette Score***.

Analysis of Factors Contributing to Employee Attrition Based on Self-Organizing Map Clustering

Abstract — ***Employee turnover can disrupt the organization's operations and more or less cause losses to the business. Therefore, it is important to understand the causal factors so that organizations can take anticipatory action. Identify reasons employees leave their jobs is crucial for both employers and policy makers, especially when the goal is to prevent this from happening. Data on the causes of employee turnover is complex data that can have many dimensions, so a certain method is needed to analyse it. In this research, an analysis of data on the causes of employee turnover with 10 dimensions will be carried out using the Self Organizing Map (SOM) method. The Self-Organizing Map (SOM) is a technique for clustering and visualizing high-dimensional data by mapping it to a two-dimensional space while preserving the data's topological structure. This neural network-based method ensures that similar data points remain close to each other in the resulting 2D representation. Self-Organizing Map will cluster the data into several uniform groups. The results of this Self-Organizing Map grouping will be assessed with the Silhouette score, Dunn index and Connectivity value to***

determine how uniform the grouping is. Hopefully that by using the results of this Self-Organizing Map grouping, it shows that the clusters formed are very good and the data is clearly grouped. Therefore, we can analyse these groups with more accurate results.

Keywords— ***Clustering; Data Analysis; Data Mining; Self-Organizing Map; Silhouette Score.***

I. INTRODUCTION

Employees are individuals recruited by a company or organization and compensated for their work. They play a crucial role in business operations, as companies rely on them to perform specific job functions. As a result, employees are among the most valuable assets that organizations must retain to optimize their contributions. Employee retention remains a significant challenge for businesses of all sizes. While research on employees' intentions to resign has a long history, the idea that resignation intent and employee commitment are influenced by distinct factors has only recently gained recognition [1]. Even though some employees remain highly satisfied and loyal to their company, it is still essential to examine the reasons behind employee resignations.

Employee attrition can be influenced by various factors [2], resulting in high-dimensional data when capturing these contributing elements. To effectively utilize this data, a data mining algorithm is required. Self-Organizing Map (SOM) is a clustering and visualization technique that projects high-dimensional data into a lower-dimensional (two-dimensional) space while preserving its topological structure [3]. As an artificial neural network, SOM groups data based on shared characteristics and highlights dominant variables within the formed clusters [4]. Cluster analysis is widely applied across different fields, including economics and employment. For instance, the SOM clustering method has been used for high-dimensional feature mapping in an analysis of 520 unplanned power plant shutdown events [5], providing deeper insights into failure reports and system performance.

Criteria for assessing goodness in determining the optimal number of clusters can be reviewed using internal validation based on the smallest connectivity index value, the largest Dunn index value, and a silhouette width value close to 1. Iwan Binanto and Andrianto Tumanggor [6] conduct research on the same dataset using k-means algorithm and silhouette coefficient to determine the best number of clusters. In contrast to them, our article research is trying to use the SOM algorithm and the smallest connectivity index value, the largest Dunn index value, and a silhouette coefficient to determine the number of clusters. A possible limitation of this research article is the data presentation which may not be suitable for some people. Because the presentation of the report on the results of data analysis is quite difficult to present to the lay public.

This study aims to analyze employee attrition by applying the Self-Organizing Map (SOM) algorithm to cluster employees based on multidimensional attributes related to resignation. The quality of the resulting clusters will be evaluated using internal validation metrics, including the Silhouette Score, Dunn Index, and Connectivity Value, to ensure the reliability of the grouping. Unlike previous studies that primarily used K-Means clustering, this research emphasizes the advantages of SOM in visualizing and interpreting complex, high-dimensional data. The ultimate goal is to uncover meaningful patterns that can assist organizations in identifying potential factors contributing to employee attrition and inform better retention strategies.

II. METHOD

This research applied SOM to analysis of factors contributing to employee attrition will be described in Figure 1. The research use google collabs with R language as a tool. This research has six steps to produce a result as an Analysis reference. The first step is preprocessing the dataset, then creating a SOM model. After that we can create clusters from the model and calculate a silhouette index, Dunn index and connectivity index value. We must conduct comprehensive evaluation from the three index values from each cluster until we define the best cluster [5]. The best cluster data will be used to analyse factors contributing to employee attrition. We will see the details from each step.

p-ISSN : 2443-2210
e-ISSN : 2443-2229

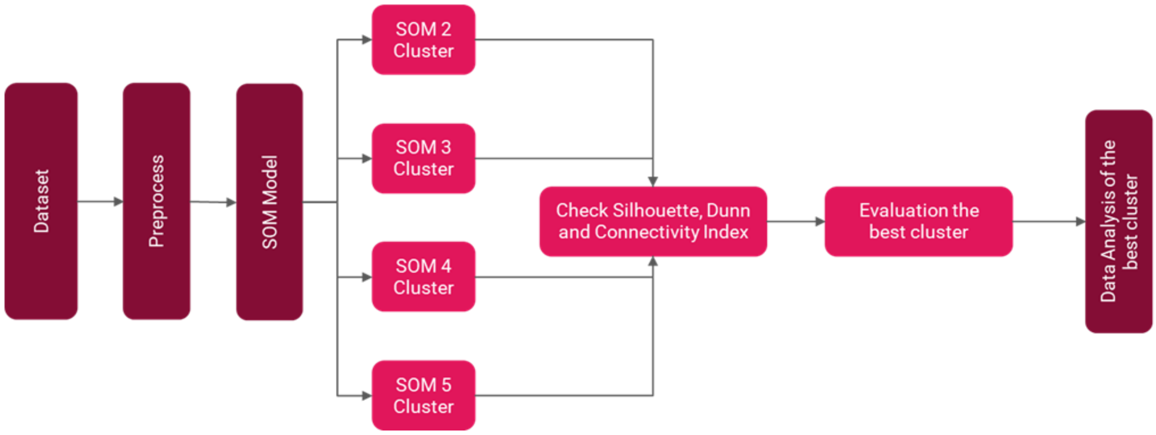

Figure 1. Proposed Method

A. Dataset

The dataset was downloaded from Kaggle. It was provided by the HR department at Salifort Motors comprises 14,999 rows and 10 columns, capturing various attributes pertaining to employee demographics, job-related factors, and potential indicators of turnover. the variables included in the dataset along with their description can be seen in Table 1.

TABLE 1
VARIABLE DATASET

No	Column Name	Description	Value
1	**satisfaction_level**	Job satisfaction level	0 – 1
2	**last_evaluation**	Score of last performance review	0 – 1
3	**number_project**	Number of projects the employee contribution	0 – 10
4	**average_monthly_hours**	Average monthly hours worked by the employee	96 – 310
5	**time_spend_company**	Length of the employee's tenure with the company (in years)	2 – 10
6	**Work_accident**	Has the employee had a workplace accident?	0 / 1
7	**Left**	Did the employee leave the company	0 / 1
8	**promotion_last_5years**	Was the employee promoted in the last 5 years	0 / 1
9	**Department**	The employee's department	String
10	**Salary**	The employee's salary	Low / medium / high

B. Data Preprocessing

This step is to clean and standardize the dataset so that the data can be processed as desired in this research. Since our research to analyse factor to employee attrition the we only use data left employee. Therefore, we have to filter the data only left=1 and delete row with left=0. We only have to change salary data from string to numeric, therefore we will convert low=1, medium=2 and high=3. The final dataset will consist 3571 rows with 9 columns and we set column "Department" as the first column. Examples of data before and after preprocessing can be seen in Figure 2.

C. SOM (self-organizing map)

The Self-Organizing Map (SOM) is a type of unsupervised artificial neural network (ANN) introduced by Teuvo Kohonen [4]. It maps high-dimensional data onto a two-dimensional grid while preserving the topological relationships within the data. Figure 3 illustrates this transformation. A key characteristic of SOMs is their ability to maintain the spatial structure of the input space, ensuring that data points that are close together in the input space remain close in the output space.

SOM operates by computing the distance (D) between input data and the weight (W) matrix, which directs the clustering process. The process starts with the creation of a SOM network map based on the provided input data. It then undergoes multiple learning iterations to refine and optimize the weight matrix [7].

$$D_j = \sum_{i=1}^{n} (w_{ij} - X_i)^2 \quad (1)$$

Where Dj is the Euclidean distance, Wij is the weight of the *i-th* neuron, Xi is the *i-th* input vector.

After getting the winning neurons, then updating the weight values of the winning neurons and neighbouring neurons using formula (2).

$$W_{ij}(t+1) = W_{ij}(t) + \alpha(t)[X_i - W_{ij}(t)] \quad (2)$$

This is a weight update rule in SOM algorithm. Each neuron (node) in SOM has a weight vector. During training, these weights are updated to better match the input data. Where W_{ij} is the weight for the *j-th* output neuron and the *i-th* input neuron, $\alpha(t)$ is the learning rate, and the neighbour function.

A data.frame: 15001 × 10

satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	left	promotion_last_5years	Department	salary
<dbl>	<dbl>	<int>	<int>	<int>	<int>	<int>	<int>	<chr>	<chr>
0.38	0.53	2	157	3	0	1	0	sales	low
0.80	0.86	5	262	6	0	1	0	sales	medium
0.11	0.88	7	272	4	0	1	0	sales	medium
0.72	0.87	5	223	5	0	1	0	sales	low
0.37	0.52	2	159	3	0	1	0	sales	low
0.41	0.50	2	153	3	0	1	0	sales	low
0.10	0.77	6	247	4	0	1	0	sales	low
0.92	0.85	5	259	5	0	1	0	sales	low
0.89	1.00	5	224	5	0	1	0	sales	low
0.42	0.53	2	142	3	0	1	0	sales	low
0.45	0.54	2	135	3	0	1	0	sales	low
0.11	0.81	6	305	4	0	1	0	sales	low
0.84	0.92	4	234	5	0	1	0	sales	low
0.41	0.55	2	148	3	0	1	0	sales	low
0.36	0.56	2	137	3	0	1	0	sales	low
0.38	0.54	2	143	3	0	1	0	sales	low
0.45	0.47	2	160	3	0	1	0	sales	low
0.78	0.99	4	255	6	0	1	0	sales	low
0.45	0.51	2	160	3	1	1	1	sales	low
0.76	0.89	5	262	5	0	1	0	sales	low
0.11	0.83	6	282	4	0	1	0	sales	low

(a)

A data.frame: 3571 × 9

Department	satisfaction_level	last_evaluation	number_project	average_montly_hours	time_spend_company	Work_accident	promotion_last_5years	salary
<chr>	<dbl>	<dbl>	<int>	<int>	<int>	<int>	<int>	<int>
sales	0.38	0.53	2	157	3	0	0	1
sales	0.80	0.86	5	262	6	0	0	2
sales	0.11	0.88	7	272	4	0	0	2
sales	0.72	0.87	5	223	5	0	0	1
sales	0.37	0.52	2	159	3	0	0	1
sales	0.41	0.50	2	153	3	0	0	1
sales	0.10	0.77	6	247	4	0	0	1
sales	0.92	0.85	5	259	5	0	0	1
sales	0.89	1.00	5	224	5	0	0	1
sales	0.42	0.53	2	142	3	0	0	1
sales	0.45	0.54	2	135	3	0	0	1
sales	0.11	0.81	6	305	4	0	0	1
sales	0.84	0.92	4	234	5	0	0	1
sales	0.41	0.55	2	148	3	0	0	1
sales	0.36	0.56	2	137	3	0	0	1

(b)

Figure 2. (a) Original dataset (b) Result pre-processed dataset

p-ISSN : 2443-2210
e-ISSN : 2443-2229

D. Internal validation

Cluster validity indices (CVIs) are broadly categorized into two main types: internal and external validation indices. External CVIs rely on external information, referred to as ground truth, to evaluate clustering results. This approach is similar to cross-validation in classification, where class labels are used to assess classification accuracy. Common examples of external validation measures include Precision, Recall, and F-measure [8].

Internal validation is validation that involves utilizing internal information to evaluate the quality of the cluster structure without relying on external information and can be used in determining the optimal number of clusters [5]. Determining the optimal number of clusters in this study uses internal validation.

1) *Silhouette Index*: The Silhouette Index is a metric used to evaluate the quality of a clustering solution for a particular data point. The value of this index ranges from -1 to +1. The best clustered data is the Silhouette Index with value closer to +1 [9]. The Silhouette Index is an unsupervised method used to assess the effectiveness of a clustering algorithm [10]. Since the Silhouette Index does not require a training set to evaluate clustering performance, it is particularly relevant to the concept of clustering. because it provides an internal evaluation of the quality of the clustering without requiring actual labels, making it ideal for assessing performance in unsupervised learning tasks such as clustering. This metric provides a way to assess the cohesion and separation of clusters based on the distance between data points.

The silhouette width value denoted by $S(i)$ can be formulated by the following equation:

$$S(i) = \frac{b_i - a_i}{max\ (b_i, a_i)}, i = 1,2,..,n \quad (3)$$

Where a_i represents the average distance of the *i-th* observation to other observations in the same cluster where:

$$a_i = \frac{1}{n\ (C(i))} \sum_{j \in C(i)} dist(i,j), \quad (4)$$

b_i represent the average minimum distance of the *i-th* observation to all other observations in the nearest neighbour cluster where:

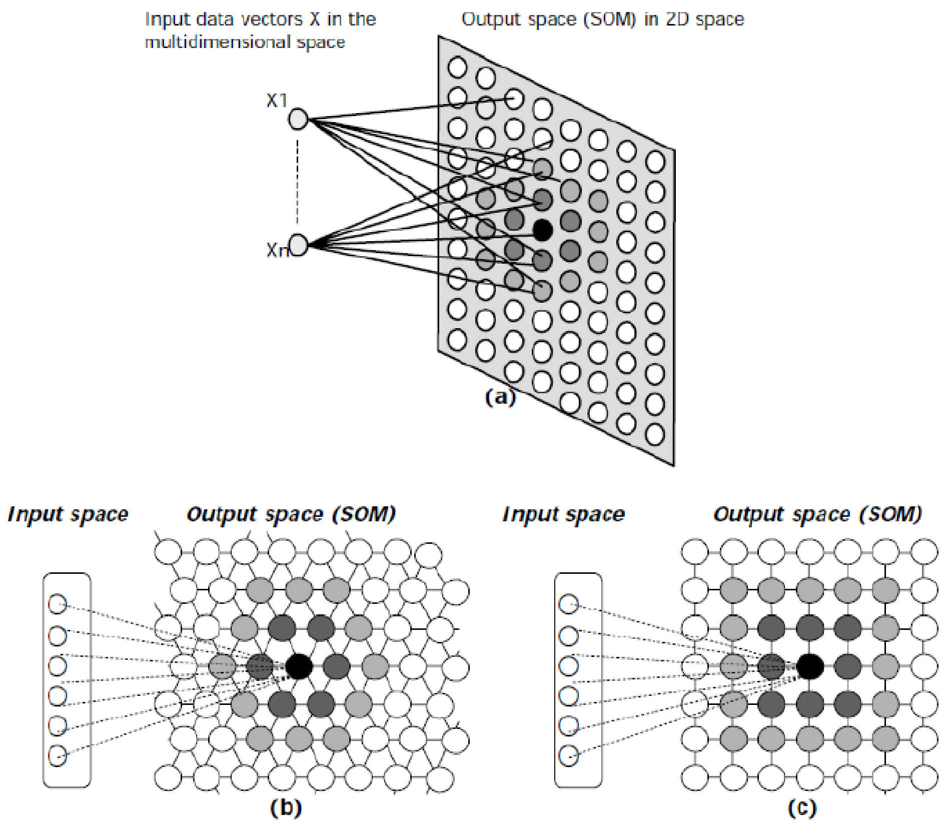

Figure 3. The structure of a SOM network, the black object indicates the node that was selected as the best match for the input pattern [11]
(a) Selection of a node and adaptation of neighbouring nodes to the input data.
(b) The hexagonal SOM grid
(c) The rectangular SOM grid

2) *Dunn Index:* Dunn's index is the ratio between the smallest distance formed between two clusters to the largest distance formed within a cluster and the higher the Dunn index value, the better [5]. The Dunn Index is calculated as the ratio of the smallest distance between observations in different clusters (inter-cluster distance) to the largest distance within a single cluster (intra-cluster distance). This index is designed to identify clusters that are well-separated and compact, making it an effective measure for evaluating the quality of a clustering solution [12].

The formula the Dunn index (DI) is:

$$DI = \frac{min_{1<i<j<k}\delta(C_i,C_j)}{max_{1\le l\le k}\Delta(C_l)} \quad (6)$$

Where $\delta(C_i, C_j)$ is the distance between cluster C_i and C_j. $\Delta(C_l)$ is diameter of cluster C_l and k is total number of clusters. The distance $\delta(C_i, C_j)$ is typically the minimum distance between any two points in cluster C_i and C_j.

$$\delta(C_i, C_j) = min\{x \in C_i, y \in C_j\} \quad (7)$$

Diameter $\Delta(C_l)$ is the maximum distance between any two points with in cluster C_l:

$$\Delta(C_l) = max\{x, y \in C_l\} \quad (8)$$

$d(x, y)$ represents the distance between point *x* and *y*.

3) *Connectivity Index:* The connectivity index is a method to show the level of cluster relationship based on the number of nearest neighbours. The value of connectivity is between zero to unlimited. The best value is the lowest in the cluster formed. The formula for the connectivity index is:

$$CI = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{|neighbors(i)|}\sum_{j\in neighbors(i)} Dist(i,j)\right) \quad (9)$$

Where *N* is the total number of units (nodes) in the SOM, i is a unit in the SOM. $|neighbors(i)|$ is the number of neighbours of node i. $neighbors(i)$ the set of neighbouring units of node i. $Dist(i,j)$ is the distance between units i and j.

III. RESULT

The best cluster we might have from the dataset, can be evaluate using internal validation. Our method uses silhouette index, Dunn Index and Connectivity value as our internal validation. We already create 2, 3, 4 and 5 cluster from SOM model to evaluate, and the result as can see in Table 2.

TABLE 2
RESULT OF SILHOUETTE INDEX, DUNN INDEX AND CONNECTIVITY VALUE FOR EACH CLUSTER

Cluster Validation	CLUSTER NO			
	2	3	4	5
silhouette	0.5261	0.5554	0.5224	0.4963
Dunn Index	0.0639	0.0059	0.1151	0.1877
Connectivity Index	187.64	187.56	177.15	199.12

The first validation is silhouette index to determine the best number of clusters because it provides a more intuitive picture of the quality of clustering. Then we verify with the Dunn Index to ensure that the clustering also has well-separated clusters and does not have very large clusters. By using these two metrics together, we can ensure that the selected clusters are not only optimal based on internal density but also have a clear separation between different clusters. Supported by connectivity index reflects the degree to which items that are near each other in the data space are also near each other in the clustering space.

Data in Table 2, give us a conclusion that Silhouette Score Indicates that 3 clusters are the best choice. Dunn Index and Connectivity Value Both show that 4 clusters are the best choice. Based on additional information from the Dunn Index and Connectivity Value, 4 clusters were selected as the best clusters because they provide good separation, good internal density, and good connectivity within the cluster.

Result clustering the dataset become 4 cluster data can be seen in Figure 4 and visualization can be seen in Figure 5.

p-ISSN : 2443-2210
e-ISSN : 2443-2229

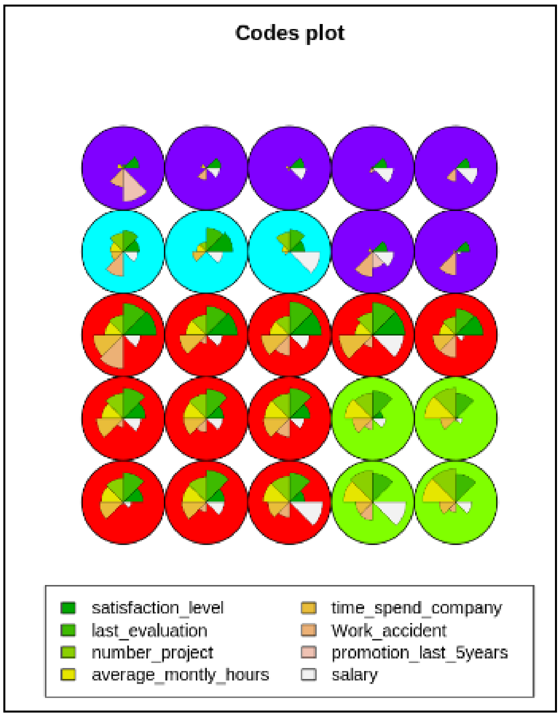

Figure 4. SOM visualization of 4 cluster in fan diagram

A data.frame: 3571 × 10

id.Department <chr>	id.satisfaction_level <dbl>	id.last_evaluation <dbl>	id.number_project <int>	id.average_montly_hours <int>	id.time_spend_company <int>	id.Work_accident <int>	id.promotion_last_5years <int>	id.salary <int>	cluster <int>
sales	0.38	0.53	2	157	3	0	0	1	3
sales	0.80	0.86	5	262	6	0	0	2	2
sales	0.11	0.88	7	272	4	0	0	2	2
sales	0.72	0.87	5	223	5	0	0	1	1
sales	0.37	0.52	2	159	3	0	0	1	3
sales	0.41	0.50	2	153	3	0	0	1	3
sales	0.10	0.77	6	247	4	0	0	1	1
sales	0.92	0.85	5	259	5	0	0	1	1
sales	0.89	1.00	5	224	5	0	0	1	1
sales	0.42	0.53	2	142	3	0	0	1	3
sales	0.45	0.54	2	135	3	0	0	1	3
sales	0.11	0.81	6	305	4	0	0	1	2
sales	0.84	0.92	4	234	5	0	0	1	1
sales	0.41	0.55	2	148	3	0	0	1	3
sales	0.36	0.56	2	137	3	0	0	1	3
sales	0.38	0.54	2	143	3	0	0	1	3
sales	0.45	0.47	2	160	3	0	0	1	3
sales	0.78	0.99	4	255	6	0	0	1	1

Figure 5. Example data result SOM Clustering of 4 cluster

Based on the visualization results of the SOM method using the hexagonal topology in Figure 4, a fan diagram is obtained which shows the grouping of employee. This study uses 8 variables with a 5x5 grid. The fan diagram has a border line to distinguish the colours in each cluster. The colours in cluster 1 are marked with a purple, cluster 2 is shown in green, cluster 3 is shown in blue and cluster 4 is shown in red. Circles represent 8 variables with different colours, the colour differences in the visualization results indicate clusters that are formed based on the variables used in the study.

The results of this clustering will be used for analysis of factors causing employee attrition. Therefore, we need to calculate and display descriptive statistics of this cluster data as we can see in Table 3.

JuTISI
Jurnal Teknik Informatika dan Sistem Informasi

TABLE 3
CALCULATION AND DISPLAY DESCRIPTIVE STATISTICS OF THE CLUSTERED DATA

Features	Cluster 1	Cluster 2	Cluster 3	Cluster 4
satisfaction_level (mean)	0.586	0.19	0.054	0.413
last_evaluation (mean)	0.89	0.857	0.788	0.524
number_project (mean)	5.02	5.9	3.94	2.14
average_montly_hours (mean)	231.26	272.72	200.88	156.74
time_spend_company (mean)	3.77 Years	3.61 Years	4.03 Years	3.01 Years
Work_accident (Percentage)	0.05%	0.04%	0.03%	0.05%
promotion_last_5years (Percentage)	0.00%	0.01%	0.00%	0.01%
salary (mean)	1.427	1.398	1.469	1.409

The interpretations that can be made are:

Cluster 1:

This cluster includes employees with relatively high satisfaction levels (0.586) and very high evaluation scores (0.89). However, their monthly working hours are also significantly high (231.26 hours), suggesting that despite being satisfied, they may be subject to overwork, which could potentially lead to burnout. The moderate tenure (3.77 years) supports this notion.

Cluster 2:

Marked by very low satisfaction (0.19) despite high evaluations (0.857), this cluster suggests a mismatch between recognition and employee feelings. Coupled with the highest number of projects (5.9) and longest working hours (272.72), it appears these employees are overburdened and underappreciated, contributing to stress and dissatisfaction.

Cluster 3:

This is the smallest cluster, possibly representing a niche group of employees with extremely low satisfaction (0.054) and low evaluation scores (0.788). Their workload is moderate, but with longer tenure (4.03 years), suggesting long-term disengagement or stagnation. Despite being few in number, this group may require targeted intervention.

Cluster 4:

The largest cluster, characterized by low satisfaction (0.413) and low evaluation (0.524), low number of projects, and lowest average monthly hours (156.74). With the shortest tenure (3.01 years), this group likely includes newer employees or those who have left early, possibly due to lack of engagement or insufficient challenge.

The number of data points in each cluster is as follows:

- Cluster 4: 1,604 data
- Cluster 1: 1,293 data
- Cluster 2: 610 data
- Cluster 3: 64 data

IV. DISCUSSION

The clustering results obtained using the Self-Organizing Map (SOM) algorithm reveal meaningful groupings of employee data based on various internal validation metrics and feature patterns. The selection of the optimal number of clusters was guided by three internal validation indices: Silhouette Score, Dunn Index, and Connectivity Index.

Cluster Validation Analysis As shown in Table 2, each cluster configuration (ranging from 2 to 5 clusters) was evaluated using three internal validation measures. The Silhouette Score peaked at 3 clusters (0.5554), indicating that, in terms of cohesion and separation, three clusters provided the clearest internal structure. However, the Dunn Index and Connectivity Index preferred the 4-cluster solution, with the highest Dunn Index (0.1877) and the lowest connectivity (177.15), indicating clearer boundaries and internal cohesiveness among the clusters.

By combining these three measures, the decision to select 4 clusters as the final configuration is justified. This balances cluster density and separation while maintaining environmental connectivity, which is important for meaningful interpretation in the context of employee behavior.

The implications for employee attrition are evident through the distinct cluster profiles identified in the analysis, which highlight key risk factors such as overwork, low satisfaction, and lack of recognition. Overwork and high expectations, as seen in Clusters 1 and 2, can lead to burnout or disengagement among employees. Furthermore, Cluster 2 highlights that a lack of appreciation despite strong performance may diminish organizational loyalty. Meanwhile, early-stage disengagement found in Cluster 4, and prolonged dissatisfaction in Cluster 3, underscore the importance of implementing early retention strategies and enhancing onboarding and employee engagement practices. By understanding these unique employee profiles, HR and management teams are better equipped to design targeted interventions, such as adjusting workloads, offering recognition, and improving career development opportunities.

V. CONCLUSION

This study applied the Self-Organizing Map (SOM) algorithm to cluster employee data and identify patterns related to attrition. Internal validation using Silhouette Score, Dunn Index, and Connectivity Index determined that four clusters offered the best balance between cohesion, separation, and neighbourhood connectivity. Each cluster revealed unique employee characteristics: Cluster 1 featured high satisfaction and evaluation but also high workload, suggesting a risk of burnout; Cluster 2 showed low satisfaction despite strong performance and heavy workload, indicating feelings of underappreciation; Cluster 3 reflected very low satisfaction and moderate workload, potentially representing long-term disengagement; while Cluster 4 consisted of employees with low involvement and short tenure, likely indicating early attrition. These insights emphasize that factors such as workload, recognition, and time spent in the company are closely linked to employee satisfaction and turnover risk, providing valuable guidance for designing more effective retention strategies.

In future work, the implementation of Deep Embedded Clustering (DEC) can be explored to improve the quality of clustering results, especially for high-dimensional employee data. Unlike traditional clustering methods, DEC integrates feature learning and clustering into a unified framework, enabling the model to automatically learn latent representations that are more suitable for clustering tasks. This approach is particularly advantageous for complex datasets where conventional methods may struggle to capture non-linear relationships. Applying DEC could reveal deeper patterns in employee behavior and attrition risk that are not easily detectable through surface-level clustering techniques like SOM or K-Means.

REFERENCES

[1] D. J. Madigan and L. E. Kim, “Towards an understanding of teacher attrition: A meta-analysis of burnout, job satisfaction, and teachers’ intentions to quit,” *Teach Teach Educ*, vol. 105, p. 103425, 2021.

[2] B. Peng, “Statistical analysis of employee retention,” in *International Conference on Statistics, Applied Mathematics, and Computing Science (CSAMCS 2021)*, K. Chen, N. Lin, R. Meštrović, T. A. Oliveira, F. Cen, and H.-M. Yin, Eds., SPIE, 2022, p. 199.

[3] P. Hartono and Y. Take, “Pairwise elastic self-organizing maps,” in *2017 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM)*, IEEE, 2017, pp. 1–7.

[4] H. Yang, D. Ma, H. Chen, and Y. Zhu, “Analysis of power plant outage event results based on SOM clustering,” *Results in Engineering*, vol. 21, p. 101995, 2024.

[5] C. Gunawan and H. Sirait, “Analysis Of Unemployment Clusters In Indonesia Using The Self Organizing MAP Method: Analysis Of Unemployment Clusters In Indonesia Using The Self Organizing MAP Method,” *International Journal of Mathematics, Statistics, and Computing (IJMSC)*, vol. 1, no. 3, pp. 8–15, 2023.

[6] I. Binanto and A. Tumanggor, “Comparison of the K-Means method with and without Principal Component Analysis (PCA) in predicting employee resignation,” in *E3S Web of Conferences 475, The International Conference on Applied Sciences and Smart Technologies (InCASST 2023)*, D. Widjaja and L. K. Budiasih, Eds., Yogyakarta: EDP Sciences, 2024.

[7] R. Widayanti, S. Madenda, E. P. Wibowo, and K. Anwar, “SOM-SIS approach to auto summary of clustering results on university academic performance,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 21, no. 1, pp. 104–112, 2023.

[8] C.-E. Ben Ncir, A. Hamza, and W. Bouaguel, “Parallel and scalable Dunn Index for the validation of big data clusters,” *Parallel Comput*, vol. 102, p. 102751, 2021.

[9] N. Nidheesh, K. A. A. Nazeer, and P. M. Ameer, “A Hierarchical Clustering algorithm based on Silhouette Index for cancer subtype discovery from genomic data,” *Neural Comput Appl*, vol. 32, no. 15, pp. 11459–11476, 2019.

[10] M. Shutaywi and N. N. Kachouie, “Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering,” *Entropy*, vol. 23, no. 759, 2021.

[11] A. Seif, M. Mokarram, and D. Sathyamoorthy, “Using Self-Organizing Maps for Alluvial Fan Classification,” *International Journal of Scientific Research in Knowledge*, vol. 2, no. 4, pp. 189–198, 2014.

[12] L. E. Ekemeyong Awong and T. Zielinska, “Comparative Analysis of the Clustering Quality in Self-Organizing Maps for Human Posture Classification,” *Sensors*, vol. 23, no. 7925, 2023.