Empowering New Capital Zones: East Kalimantan's Economic District Outlooks Using Location Quotient and Cluster Analysis

Mega Silfiani, Diana Nurlaily and Irma Fitria

***Abstract*—This research focuses on investigating the economy of the new capital buffer zone by identifying and clustering its leading sectors in GRDP (Gross Regional Domestic Product) of East Kalimantan. The identification of a region's leading sector through LQ (Location Quotient) index has proven to be effective. In addition, k-means clustering and Self-Organizing Maps (SOM) are adopted to provide comprehensive insights. The results show that LQ index quickly identifies the main sectors in each district of East Kalimantan. In addition, the k-means clustering has better performance than SOM based on the Silhouette coefficient. This meticulous analysis confirms the existence of two distinct clusters, one including eight members and the other consisting of only two. Anticipating future research endeavours, the exploration of various approaches for constructing clusters, encompassing both hierarchical and non-hierarchical approaches, provides the potential to enhance the performance of clusters. By investigating this structure, a more comprehensive comprehension of the economic framework of East Kalimantan can be achieved, as well as its potential role as a buffer for the capital region.**

***Keywords*—GRDP, K-Means, Location Quotient, Self-Organizing Map, Silhouette Coefficient.**

I. INTRODUCTION

THE relocation of the capital of Indonesia "Jakarta" to the "Nusantara" in East Kalimantan will contribute to national economic development by 0.1%, reflected by a rise in real gross domestic product (GDP) [1]. The growth corresponds to the utilization of previously unexploited potential resources, such as acquiring land for productive infrastructure demands and developing employment opportunities for skilled human resources [2]. This action will also have a direct effect on Kalimantan, the lungs of the world and a biodiversity hotspot. If development preparation and execution are unreliable, deforestation for oil palm plantations and industrial operations will have a greater impact on the environment. Relocating the capital city will not only have a negative impact on the environment, but also on the social life of residents in the new capital buffer zone if the educational and professional backgrounds of the local community remain stagnant. This leads to socioeconomic and health disparities between the existing population and migrants. To prevent social conflict, governments ought to anticipate responses to environmental, social, and economic issues.

M. Silfiani, D. Nurlaily, I. Fitria are with the Department of Statistics, Institut Teknologi Kalimantan, Balikpapan 76126, Indonesia e-mail: megasilfiani@lecturer.itk.ac.id

Manuscript received July 18, 2023; accepted September 14, 2023.

A potential strategy for addressing this issue involves the identification and clustering of significant sectors of the Gross Regional Domestic Product (GRDP) that have the potential to foster economic expansion within the newly established capital buffer zones. The Gross Regional Domestic Product (GRDP) is a macroeconomic indicator that can be considered an indicator of productivity that reflects the total value of products and services produced by a region each year. GRDP has a significant and positive effect on a region's human development index (HDI) [3], [4]. HDI is an indicator that refers to how residents of a region can utilize the results of a development as part of their rights to income, health, education, etc. Consequently, it is essential for local administrations to be able to grow GRDP in order to enhance welfare. LQ (Location Quotient) index is a useful tool to identify a region's dominant sectors. This index is characterized by its simplicity and convenience, as well as its straightforward objective in selecting appropriate indices [5], [6], [7], [8]. However, the current limitations do not allow for a conclusive determination regarding the strategic sectors that have been identified [5]. Therefore, we require cluster analysis for more investigations that can classify these prominent sectors. K-means clustering, and Self-Organizing Map are two widely used clustering methods. The k-means clustering is easier to implement and faster than most other clustering algorithms [9], [10]. It also can be applied in both large and small data sets and produce the best clustering if the generated cluster has a converging feature [10]. K-means clustering is applied to various fields, such as pattern recognition, natural language processing, and medical diagnosis [10]. In addition, K-means clustering was also adopted for region clustering such as based on power supply system [11] and natural soil environment [12]. Meanwhile, SOM is a type of artificial neural network trained by unsupervised learning to classify and cluster data without interpreter bias [13]. It gains by being able to use mixed type feature data, e.g., numerical and categorical features [14]. The implementation of SOM includes various domains, such as text mining [15], learning behavior [16], and region mapping [17]. The comparative investigations undertaken in prior studies [11], [12], [13], [14], [15], [16], [17] have revealed a noteworthy divergence in outcomes when assessing the efficacy of k-means and self-organizing map (SOM) techniques across various scenarios, thereby prompting a fundamental query regarding their performance in the clustering of the GRDP sectors within East Kalimantan province.

Therefore, this study aims to identify and to cluster economic potential through sectoral Gross Regional Domestic Product (GRDP) in buffer zones of new capital of Nusantara. Method for identification of leading sectors in GDP using LQ index. Meanwhile, to cluster regions based on sectors in GRDP, a comparison of k-means and SOM methods is used to get the best cluster. This study is anticipated to offer valuable insights that will assist governments to reach rational policy regarding to resource allocation and policies associated to economic development. By obtaining understanding of the primary sectors contributing to the Gross Regional Domestic Product (GRDP) in each region, the government can strategically allocate investments and aid the sectors that exhibit the highest potential for fostering economic growth. Furthermore, this study can enhance public consciousness of the economic prospects and opportunities inside their respective localities. Additionally, it may serve as a catalyst for local entrepreneurial endeavours aimed at mitigating regional economic disparities.

The remainder of the paper is organized as follows. Section II explains the dataset and review's location quotient index, k-means clustering, self-organizing map, and silhouette coefficient in detail. In sec:3, we analyze data using the proposed methods. The k-means clustering, and self-organizing map are com-pared under silhouette coefficient using datasets in sec:2. Finally, the paper concludes in sec:4 with suggested future work.

II. Methodology

A. Dataset

This research relies on secondary data sourced from the Badan Pusat Statistik (Central Bureau of Statistics). The dataset encompasses key economic indicators for the year 2022, focusing on the East Kalimantan Province. East Kalimantan Province consists of ten cities, i.e., Berau, Kutai Barat, Katai Kartanegara, Kutai Timur, Mahakam Ulu, Paser, Penajam Paser Utara (PPU), Balikpapan, Bontang and Samarinda. The ten cities are illustrated by fig:1.

The dataset comprises the Gross Regional Domestic Product (GRDP) at con-stant price categorized by business sectors and city within East Kalimantan. Ad-ditionally, it includes GRDP at constant prices categorized according to the busi-ness sector of East Kalimantan Province. The specific sectors constituting the GRDP breakdown are outlined in tab:1.

B. Location Quotient Index

Regional economies exhibit distinctive characteristics in terms of their economic interdependencies, which set them apart from national economies [8]. A notable illustration lies in the presence of intermediate inputs originating from multiple locations within a given country. While this phenomenon may result in a reduction of economic activity at the local level, it is nevertheless considered as domestic output when viewed from a national perspective. Overlooking these nuances could potentially lead to an overestimation of the value attributed to specific industries. Consequently, to address this challenge and meet the demand for accurate regional economic models, several indirect estimation techniques have been proposed. One practical and easily applicable approach to localizing

Fig. 1: East Kalimantan Province Map
Source: Kalimantan Timur (2023) [18]

TABLE I: Gross Regional Domestic Product Based on Business Field Sector

Variable	Business Field Sector
X_1	Agriculture, Forestry and Fisheries
X_2	Mining and Quarrying
X_3	Processing Industry
X_4	Procurement of Electricity and Gas
X_5	Water Supply, Waste Management, Waste and Recycling
X_6	Construction
X_7	Wholesale and Retail Trade; Car and Motorcycle Repair
X_8	Transportation and Warehousing
X_9	Provision of accommodation and food and drink
X_{10}	Information and Communication
X_{11}	Financial Services
X_{12}	Real Estate
X_{13}	Company Services
X_{14}	Government Administration, Defense and Mandatory Social Security
X_{15}	Educational Services
X_{16}	Health and Social Activities Services
X_{17}	Other Services

a national input/output table involves the implementation of a series of Location Quotients (LQs) index based on Gross Regional Domestic Product (GRDP) [8].

The concept of location quotient was initially introduced by P. Haggett and has since been extensively employed in the field of location analysis, particularly in studies pertaining to industrial and economic location research [18]. Location quotient (LQ) index employs the leading or basis sector to identify economic sector focus [5]. The location quotient is the ratio of the sector i output share in the city or district to the sector i output share in the province. Here, the leading sector refers to the business sector that can be exploited constantly by the regional government. Location quotient is a tool for economic development that has positive and negative aspects [6]. The fundamental economic models identify development sectors using the LQ index. On the contrary, LQ index assesses the relative concentration or degree of a focus of the economy. LQ index evaluates economic conditions, defines the specialization of economic activity, or assesses relative concentration in order to determine which sectors dominate industrial economic activity [5]. The labor and income aspects are the focus of the LQ index. The LQ approach is incapable of producing conclusive findings regarding the emphasized strategic sectors. In the preliminary stage, it is sufficient to provide an overview of a region's strength in a specific sector. The following formula provides for comparing the capacities of sectors in a region [5]:

$$\mathrm{LQ} = \frac{V_i}{\sum_I V_i} \bigg/ \frac{y_i}{\sum_I y_i} \tag{1}$$

where v_i is the GRDP of i-sector at a lower regional level, $\sum_I V_i$ is total GRDP at a lower regional level, y_i is the GRDP of i-sector at a higher regional level and $\sum_I y_i$ is a total GRDP at a higher regional level.

C. K-Means Clustering

The k-means clustering is a common hierarchical clustering algorithm that is simple to use [9]. It is multivariate analysis that often applied in data mining method and is capable of organizing data into one or more clusters such that data with similar characteristics are arranged together and data with distinct features are organized into other groups [7]. K-Means is a distance-based clustering technique that divides data into multiple clusters and can only be applied to numerical characteristics. The following are the phases of k-means clustering:

1) Determine k as the number of clusters.
2) Determine the random value to be used as the initial cluster centroid for k cluster. Calculate the distance of each observation data to each centroid using the Euclidean distance formula as follows.

$$d(x_i, \mu_j) = \sqrt{\sum_{k=1}^{p} (x_{ik} - \mu_{jk}^2} \tag{2}$$

where $d(x_i, \mu_j)$ is the Euclidean distance between the i-th GDRP sector and the j-th cluster centroid, x_{ik} is GRDP sector at i-th, is the centroid at j-th and p is data dimension
3) Grouping each data based on its proximity to the centroid or determine the smallest distance
4) Recalculating the j-th cluster centroid using the following equation

$$\mu_{jk} = \frac{1}{n_j} \sum_{j=i}^{n_j} x_{ik} \tag{3}$$

where nj is a number of data points assigned to cluster j.
5) If the data for each cluster has not stopped, then repeat from steps 2 to 5 until nothing changes in the members of each cluster.

D. Self-Organizing Map (SOM)

Self-Organizing Map (SOM) introduced by Teuvo Kohonen in 1982 is an unsupervised based on cognitive learning [19]. It represents a multidimensional dataset in a two-dimensional visualization which can be understood and properly interpreted [10]. It is a network that does not require special supervision. The term Maps indicates that this procedure weights input data using maps. Each node in the SOM network attempts to resemble every input that the network has received [8]. The structure of a self-organizing map (SOM) comprises an input layer and an output layer that exhibit extensive interconnections, with each connection being assigned a weight [20]. The output is structured in a two-dimensional arrangement of neurons, wherein the display of the output data is incorporated throughout the learning process [20]. The steps of clustering in the SOM are as follows:

1) Initialization
   a) Initialize 2D grid of neurons with random initial weights. Each neuron is represented by weight vector wij for neuron i in the grid and dimension j.
   b) Initialize learning rate (η) and neighborhood size (σ). The common learning rate is 0.1, 0.01, and 0.001. Meanwhile, neighborhood size is initialized by relatively a large value.
2) Training
   a) Competition. Calculate the Euclidian distance (dij) between input vector x and each neuron's weight vector (w_{ijk}) which is calculated by the following equation:

$$d_{ij} = \sqrt{\sum_{k=1}^{D} (x_k - w_{ijk})^2} \tag{4}$$

   b) Cooperation. After the distance between nodes is obtained, the minimum value is determined from the calculation of the distance vector d_{ij}, then the next step is to update the weights.

$$w_{ijk}(t+1) = w_{ijk}(t) + \eta(t) h_{ij}(t) \left[x_k - w_{ijk}(t) \right] \tag{5}$$

   where $w_{ijk}(t)$ is the weight for neuron j in dimension j at time t, $\eta(t)$ is the learning rate at time t, and $h_{ij}(t)$ is neighborhood function typically a Gaussian function of neighborhood size.
   c) In the process of obtaining a new weight, learning rate (η) and neighborhood size (σ) usually decrease over time.
   d) The condition for stopping the test is done by calculating the difference between the $w_{ijk}(t+1)$ and $w_{ijk}(t)$. If the w_{ijk} value changes slightly, it means that the test has converged so that it can be stopped.

E. Silhouette Coefficient

The assessment of clustering models is a crucial task in various research do-mains, aiming to determine the efficacy of data grouping techniques [9]. Silhouette coefficient is one of the effective statistical ways for determining the cluster count value [10]. Among the methods employed for evaluating the quality of clustering outcomes, the silhouette coefficient has gained prominence. By utilizing this metric, researchers can gain insights into the degree of separation be-tween clusters and the compactness of individual clusters. The following are the steps in calculating the silhouette coefficient [9].

1) Compute the average dissimilarity within a cluster.

$$a(i) = \frac{\sum_{j \in C_{i+j}} d(i, j)}{|C_i| - 1} \tag{6}$$

where $a(i)$ represents the average dissimilarity of data point i within its cluster, $|C_i|$ is the total number of data points in the cluster, and $d((i, j)$ denotes the dissimilarity between data points i and j.

2) Calculate the average dissimilarity to neighboring clusters.

$$b(i) = \min C_{k,k \neq i} \left\{ \frac{\sum_{k \in C_k} d(i,k)}{|C_k|} \right\} \tag{7}$$

where $b(i)$ represents the average dissimilarity of data point i to the neighboring cluster, $|C_k|$ is the total number of data points in the neighboring cluster, and $d(i,k)$ represents the dissimilarity between data points i and k.

3) Compute the silhouette coefficient. The silhouette coefficient for each data point is calculated by subtracting the average dissimilarity within the cluster from the average dissimilarity to the neighboring cluster, and then dividing the result by the maximum of these two values. This can be represented as:

$$y(j) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{8}$$

III. Result and Discussion

A. Location quotient index

Location quotient index is used to identify the leading sector of Gross Regional Domestic Product (GRDP) for each district in East Kalimantan Province. The number of sectors in GRDP in this research are 17 sectors, with detailed information based on tab:1. The location quotient (LQ) index for each city in East Kalimantan Province is calculated based on business sectoral GRDP for each city. If LQ index greater than one indicates that a specific sector holds a higher significance level within the local economy than its overall importance in the national economy. This demonstrates the potential for the region to emerge as a front runner in its respective industry. LQ values below one signifies that the sector's significance to the local economy is comparatively lower than its significance to the national economy. tab:2 shows that the leading sector most owned by districts in the province of East Kalimantan is (X_1) Agriculture, Forestry, and Fisheries, the second most is (X_2) Mining and Quarrying. In contrast, the smallest leading sectors are (X_3) Processing Industry, (X_{10}) Information and Communication and (X_{11}) Financial Services. In addition, if based on districts, the most districts that have a leading sector is Samarinda, the second is Balikpapan. In contrast, Kutai Timur is the district that has the least leading sector and is followed by Mahakam Ulu.

B. K-Means Clustering

The process of creating clusters using the k-means clustering begins with deter-mining the desired number of clusters (k). In this study, the optimal number of clusters is determined by evaluating the best results for cluster numbers ranging from $k = 2$ to $k = 9$. The evaluation of these results is based on the average silhouette width value for each cluster number from $k = 2$ to $k = 9$. After conducting experiments involving the creation of 2 to 9 clusters, the average silhouette width values were compared to identify the highest value. A higher average silhouette width value closer to 1 indicates better cluster formation for grouping GRDP per sector. Average silhouette coefficient is the arithmetic mean of vector of silhouette coefficient for each city. Meanwhile, silhouette coefficient is calculating based on 8. As shown in tab:3, the highest overall average silhouette width value obtained is 0.5768, which occurs when the optimal number of clusters (k) is set to 2. Therefore, it can be inferred that the k-means clustering generates the optimum number of clusters, which is 2 clusters.

C. Self-Organizing Map (SOM)

In this study, the SOM method was utilized to implement unsupervised learning using a two-dimensional approach. Accordingly, a two-dimensional grid topology was established as the foundational framework. Specifically, a hexagonal grid topology was employed, limited to a maximum of 10 grids. Within the SOM methodology, input vectors are assigned to respective grids based on the weights connecting the input and output neurons. To initiate the clustering process, initial weights (w_{ijk}) linking the input and output neurons were randomly assigned values ranging from 0 to 1. Subsequently, a random input vector, represented as x, was selected and introduced to the input neurons. The resulting clustering outcomes for varying grid sizes were evaluated by calculating the average silhouette width value, as illustrated in Table 4. Average silhouette coefficient is the arithmetic mean of vector of silhouette coefficient for each city. Meanwhile, silhouette coefficient is calculating based on 8. Hence, before calculating the silhouette coefficient, we must obtain the information about which grid unit (or neuron) in the SOM each data point has been assigned to.

The criteria for determining the optimal cluster in the SOM method aligns with that of the k-means clustering, relying on the average silhouette width value. Through the application of the SOM method, the highest average silhouette width value was obtained for a grid size of 2×1, measuring 0.5399. Thus, it can be inferred that the SOM method generates the optimal number of clusters, resulting in 2 distinct clusters.

D. Comparison of K-Means Clustering and Self-Organizing Map (SOM)

The previous sub-sections presented the process of cluster formation using the k-means clustering and SOM methods. Additionally, the evaluation of these clustering methods was performed, focusing on the average silhouette width value. The subsequent section provides the highest average silhouette

TABLE II: Location Quotient Index Based on Business Sectors and Cities

Cities	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	$X12$	X_{13}	$X14$	X_{15}	X_{16}	X_{17}
Berau	1.37	**1.34**	0.19	0.65	0.83	0.48	**1.08**	**1.68**	**1.07**	0.66	0.31	**1.00**	0.48	0.57	**1.72**	**1.68**	**1.15**
Kutai Barat	**1.75**	**1.18**	0.26	0.45	0.68	**1.25**	**1.22**	0.55	0.33	0.81	0.08	0.63	0.28	**2.29**	**1.04**	**1.19**	0.51
Kutai Kartanegara	**1.59**	**1.48**	0.17	0.98	0.73	0.84	0.60	0.34	0.28	0.53	0.20	0.52	0.14	0.78	0.68	**1.09**	0.38
Kutai Timur	0.94	**1.77**	0.14	0.22	0.26	0.26	0.34	0.35	0.28	0.19	0.11	0.42	0.34	0.45	0.66	0.24	0.29
Mahakam Ulu	**11.23**	0.14	0.03	0.56	0.78	0.82	0.80	**1.36**	0.14	0.19	0.05	0.24	0.17	0.74	0.12	0.19	0.10
Paser	**1.58**	**1.57**	0.23	0.41	0.41	0.30	0.67	0.15	0.30	0.49	0.24	0.35	0.25	0.59	0.87	0.73	0.35
PPU	**2.71**	0.52	0.65	**1.39**	**1.78**	**2.86**	**1.52**	0.55	0.48	0.94	0.47	**1.43**	0.08	**1.88**	**2.15**	0.04	**1.43**
Balikpapan	0.13	0.00	**2.86**	**1.46**	**1.36**	**1.57**	**1.45**	**2.33**	**1.64**	**2.16**	**2.08**	**1.81**	**1.31**	0.66	0.89	0.88	**1.18**
Bontang	0.17	0.01	**4.06**	0.66	0.51	0.77	0.54	0.58	0.60	0.59	0.55	0.61	2.77	0.71	0.73	1.16	0.64
Samarinda	0.23	0.28	0.37	**2.88**	**3.49**	**2.45**	**2.93**	**2.36**	**4.36**	**2.73**	**4.74**	**2.83**	**3.87**	**3.40**	**2.44**	**2.22**	**4.56**

TABLE III: Average silhouette width for k-means clustering

Number of cluster	Average silhouette width
2	0.5768
3	0.5020
4	0.4350
5	0.2024
6	0.2167
7	0.1913
8	0.1532
9	0.0892

TABLE IV: Average silhouette width for Self Organizing Map (SOM)

Grid size	Number of cluster	Average silhouette width
2x1	2	0.5399
2x2	4	0.2858
2x3	6	0.2390
2x4	8	0.1973
2x5	10	0.0812
3x1	3	0.5020
3x2	6	0.2205
3x3	9	0.2205
4x1	4	0.1905
4x2	8	0.2187
5x1	5	0.1670
5x2	10	0.2187

width value achieved after generating clusters using both the k-means clustering and SOM methods.

According to the data presented in tab:5, the k-means clustering yielded the highest average silhouette width value of 0.5768. This value signifies better cluster formation. Furthermore, the optimal number of clusters generated by the k-means clustering was determined to be 2 clusters. Consequently, it can be concluded that the k-means clustering is the preferred clustering approach for deter-mining GRDP sector cluster.

E. Analizing of Cluster

After obtaining the optimum number of clusters according to the silhouette average width, we can describe the position of the district according to the k-means cluster. There are ten districts in the province of East Kalimantan, namely (1) Berau, (2) Kutai Barat, (3) Kutai Kartanegara, (4) Kutai Timur, (5) Mahakam Ulu, (6) Paser, (7) PPU, (8) Balikpapan, (9) Bontang and (10) Samarinda.

TABLE V: Comparison of Average silhouette width for K-means and SOM

Method	Number of cluster	Average silhouette width
K-means	2	0.5768
Self Organizing Map	2	0.5399

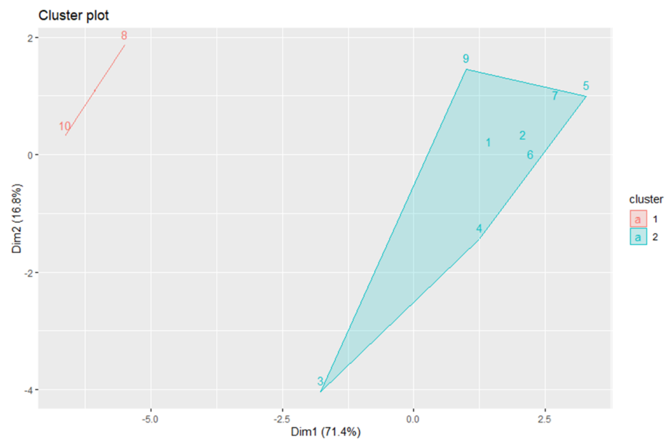

Fig. 2: Cluster plot based on k-means for East Kalimantan District

We illustrate the clusters using $f_{vizcluster}$ in R which shows the clusters using the first two principal components to define the X-Y coordinates of each city. fig:2 explains that there are 2 clusters. The first cluster consists of 2 districts i.e., Samarinda and Balikpapan, while the second cluster consists of 8 districts i.e., Berau, West Kutai, Kutai Kartanegara, East Kutai, Mahakam Ulu, Paser and PPU. For a more comprehensive understanding of the district grouping in East Kalimantan based on the GRDP sector, it is valuable to examine the cluster results depicted in fig:2 alongside the outcomes of the location quotient index illustrated in tab:1. fig:3 is LQ index based on k-means clustering for each business sectoral.

fig:3 can be deduced that the first cluster, consisting of Samarinda and Balikpapan, exhibits the highest number of districts with prominent sectors. The biggest 3 of leading sectors for first cluster are Provision of accommodation and food and drink (X_{11}), Financial Services (X_9), and Other services (X_{17}). Notably, the cluster members in this group do not heavily rely on the GRDP inputs from Agriculture, Forestry, and Fisheries (X_1) and Mining and Quarrying (X_2). Conversely, the second cluster, consisting of eight districts, predominantly shares either the leading sector Agriculture, Forestry, and Fisheries (X_1) and Mining and Quarrying (X_2) in terms of GRDP, except for Bontang.

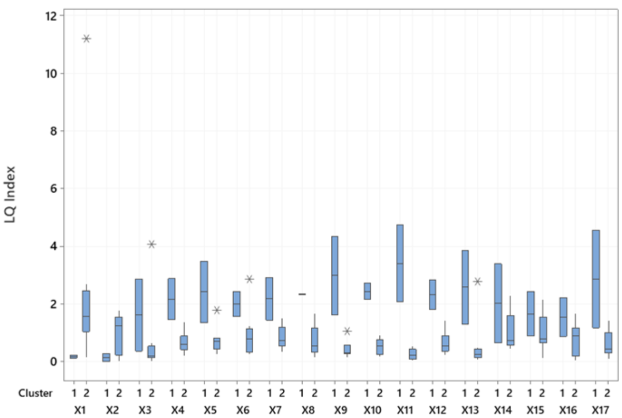

Fig. 3: LQ index based on k-means clustering

IV. Conclusion

This study aims to enhance the economic prospects of the new capital of Nusantara buffer zone by identifying and clustering key of Gross Regional Domestic Product (GRDP) sectors using LQ index, k-means clustering, and SOM. The findings from LQ index reveal the dominant leading sectors in each district of East Kalimantan province i.e., (X_1) Agriculture, Forestry, and Fisheries, the second most is (X_2) Mining and Quarrying. Meanwhile, the k-means clustering, and SOM produce each two clusters. However, based on the Silhouette Coefficient, the k-means clustering outperforms SOM. The resulting clusters from k-means clustering exhibit eight members in one cluster and two members in the other. In addition, LQ index based on cluster explain that first cluster has top 3 leading sector in are Provision of accommodation and food and drink (X_{11}), Financial Services (X_9), and Other services (X_{17}). Meanwhile, the second cluster has biggest leading sector in Agriculture, Forestry, and Fisheries (X_1) and Mining and Quarrying (X_2). Future research endeavors may explore alternative clustering methods, encompassing hierarchical and non-hierarchical approaches, to enrich the analysis.

References

[1] M. P. P. Nasional and K. Bappenas, "Dampak ekonomi dan skema pembiayaan pemindahan ibu kota negara," *Disampaikan dalam Dialog Nasional II Pemindahan Ibu Kota Negara, Menuju Ibu Kota Masa Depan: Smart, Green and Beautiful, Jakarta*, vol. 26, 2019.

[2] R. R. A. Hasibuan and S. Aisa, "Dampak dan resiko perpindahan ibu kota terhadap ekonomi di indonesia," *AT-TAWASSUTH: Jurnal Ekonomi Islam*, vol. 5, no. 1, pp. 183–203, 2020.

[3] M. Ariwuni and I. Kartika, "Pengaruh pdrb dan pengeluaran pemerintah terhadap ipm dan tingkat kemiskinan di kabupaten/kota provinsi bali," *E-Jurnal Ekonomi Pembangunan*, vol. 8, no. 12, pp. 2807–3114, 2019.

[4] M. Muliza, T. Zulham, and C. Seftarita, "Analisis pengaruh belanja pendidikan, belanja kesehatan, tingkat kemiskinan dan pdrb terhadap ipm di provinsi aceh," *Jurnal Perspektif Ekonomi Darussalam (Darussalam Journal of Economic Perspec*, vol. 3, no. 1, pp. 51–69, 2017.

[5] K. R. Jumiyanti, "Analisis location quotient dalam penentuan sektor basis dan non basis di kabupaten gorontalo," *Gorontalo Development Review*, vol. 1, no. 1, pp. 29–43, 2018.

[6] F. Dong, Y. Wang, L. Zheng, J. Li, and S. Xie, "Can industrial agglomeration promote pollution agglomeration? evidence from china," *Journal of Cleaner Production*, vol. 246, p. 118960, 2020.

[7] G. Yin, Z. Lin, X. Jiang, M. Qiu, and J. Sun, "How do the industrial land use intensity and dominant industries guide the urban land use? evidences from 19 industrial land categories in ten cities of china," *Sustainable Cities and Society*, vol. 53, p. 101978, 2020.

[8] K. Morrissey, "Producing regional production multipliers for irish marine sector policy: A location quotient approach," *Ocean & coastal management*, vol. 91, pp. 58–64, 2014.

[9] Ž. D. Vlaović, B. L. Stepanov, A. S. Anelković, V. M. Rajs, Z. M. Čepić, and M. A. Tomić, "Mapping energy sustainability using the kohonen self-organizing maps-case study," *Journal of Cleaner Production*, vol. 412, p. 137351, 2023.

[10] A. Abdulnassar and L. R. Nair, "Performance analysis of kmeans with modified initial centroid selection algorithms and developed kmeans9+ model," *Measurement: Sensors*, vol. 25, p. 100666, 2023.

[11] G. Niu, Y. Ji, Z. Zhang, W. Wang, J. Chen, and P. Yu, "Clustering analysis of typical scenarios of island power supply system by using cohesive hierarchical clustering based k-means clustering method," *Energy Reports*, vol. 7, pp. 250–256, 2021.

[12] W. Zhao, J. Ma, Q. Liu, J. Song, M. Tysklind, C. Liu, D. Wang, Y. Qu, Y. Wu, and F. Wu, "Comparison and application of sofm, fuzzy c-means and k-means clustering algorithms for natural soil environment regionalization in china," *Environmental Research*, vol. 216, p. 114519, 2023.

[13] H. Bedle, C. R. Garneau, and A. Vera-Arroyo, "Clustering energy support beliefs to reveal unique sub-populations using self-organizing maps," *Heliyon*, vol. 9, no. 7, 2023.

[14] S. Ayesha, M. K. Hanif, and R. Talib, "Overview and comparative study of dimensionality reduction techniques for high dimensional data," *Information Fusion*, vol. 59, pp. 44–58, 2020.

[15] H.-C. Yang, C.-H. Lee, and H.-W. Hsiao, "Incorporating self-organizing map with text mining techniques for text hierarchy generation," *Applied Soft Computing*, vol. 34, pp. 251–259, 2015.

[16] N. B. Ahmad, U. F. Alias, N. Mohamad, and N. Yusof, "Principal component analysis and self-organizing map clustering for student browsing behaviour analysis," *Procedia Computer Science*, vol. 163, pp. 550–559, 2019.

[17] P. Zeng, F. Sun, Y. Liu, Y. Wang, G. Li, and Y. Che, "Mapping future droughts under global warming across china: A combined multi-timescale meteorological drought index and som-kmeans approach," *Weather and Climate Extremes*, vol. 31, p. 100304, 2021.

[18] S. B. Billings and E. B. Johnson, "The location quotient as an estimator of industrial concentration," *Regional Science and Urban Economics*, vol. 42, no. 4, pp. 642–647, 2012.

[19] R. Z. Cabada, M. L. B. Estrada, and C. A. R. García, "Educa: A web 2.0 authoring tool for developing adaptive and intelligent tutoring systems using a kohonen network," *Expert Systems with Applications*, vol. 38, no. 8, pp. 9522–9529, 2011.

[20] A. A. Hameed, B. Karlik, M. S. Salman, and G. Eleyan, "Robust adaptive learning approach to self-organizing maps," *Knowledge-Based Systems*, vol. 171, pp. 25–36, 2019.