Kasetsart Journal of Social Sciences 42 (2021) 463–472

Kasetsart Journal of Social Sciences

journal homepage: http://kjss.kasetsart.org

The development of special economic zone and the cluster approaches: The case in Southern Thailand

Arisara Romyen

Department of Agribusiness Economics and Management, Faculty of Economics, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand

Article Info

Article history:
Received 20 March 2020
Revised 10 June 2020
Accepted 16 June 2020
Available online 31 July 2021

Keywords:
clustering validation measurements,
hierarchical agglomerative clustering,
k-means clustering algorithm,
model-based clustering,
Songkhla Special Economic Zone

Abstract

A Special Economic Zone means a key role in the development of regions. This raises the question of whether production sectors can be designed to facilitate clustering. In this paper, we determine the SEZ in Southern Thailand. For the clustering analyses, the K-means algorithm, the Hierarchical clustering algorithm and the Model-Based clustering algorithm were applied as well as clustering validation measurements being carried out. Moreover, the Silhouette width (0.73), and Dunn index (for 4 cluster = 0.983 against 5 cluster = 0.232) were executed to ensure the clustering validation. Consequently, the four optimal clusters, consisting of cluster 1 (AGRI, SALE), cluster 2 (ELEC, HOTE, TRAN, FINAN, ESTE), cluster 3 (FISH, CONS), and cluster 4 (MANU, MIN), were specified. These classified tier clusters explicitly demonstrate current economic performance at Songkhla's SEZ. The policy-maker can adopt this clustering scheme to gain advantages through SEZ.

Introduction

An aspect of dense economic activity becomes an essential criterion especially for integrated economies. The existence of agglomerative economies leads to productivity advantages that emerge from production sectors concentrated relative to one another. The cogent evidence of advantages relates to agglomeration guides about spatial production policies that influence the performance of firms to enhance the pace of production (World Bank, 2009).

In the current context of globalization, several developing countries have been implementing the Special Economic Zone (SEZ) to promote a market-led economy and attract Foreign Direct Investment (FDI). From 1961 onwards, Thailand has adopted the SEZ development policies in which the beginning stage emphasizes backbone policies such as national development planning, key infrastructure projects, promoting regional and rural development. In 2014, the Thai government commenced the establishment of SEZ pilot areas covering five border areas, namely Tak, Mukdahan, Sa Kaeo, Trat, and Songkhla. In this paper, we examine the specific instrument of spatial economic policy to be establishing-well functioning Special Economic Zones.

For a deep look inside Songkhla's SEZ, it has been designed to be for processing industries for export and multimodal transport. The target activities include; (1) agricultural, fishery industries, (2) petrochemical industry, (3) manufacturing-seafood processing industry, rubber-

E-mail address: arisara.r@psu.ac.th.

https://doi.org/10.34044/j.kjss.2021.42.3.03
2542-3151/

processed products, (4) multimodal transport, (5) industrial estate/zone, and (6) tourism-related industry (Cherdchom, Prommin, & Romyen 2016). At present, Sadao and Padang Besar International Border Checkpoints at Songkhla are firstly ranked in terms of the cross-border trading volume of Thailand (National Economic and Social Development Board [NESD], 2013). As a consequence, Songkhla's SEZ is deemed as a considerable potential for growth. Romyen, Liu, Sriboonchitta, and Cherdchom, (2019) verified that foreign direct investment (FDI) and border trade significantly contribute towards Songkhla's SEZ. Somehow, their economic performances such as the regional economic growth and FDI have sharp fluctuations, but the border trade manifests constantly. Hence, the government should enhance its competitiveness of FDI and trade policies.

Recent works on regional development guide that production agglomerations can impart potential competitiveness since this concept pulls in more customers and suppliers rather than what a single firm can gain alone (Becattini et al., 2003; Cossentino et al., 1996; Porter, 2000). Moreover, clusters facilitate the flexible management of production with high levels of leverage and specialization (Propris & Driffield, 2005). Several studies on Thailand widely examine the general determinants of increase in FDI; however, studies on clustering analysis of the SEZ development are seldom in literature. Therefore, it is essential to ascertain the influences of each production. Thereby, the production of goods and services in Songkhla's SEZ might be classified for determining its potential competitiveness in a global market.

This study aims to investigate the regional distribution of the SEZ policies in Southern Thailand using cluster analysis to encourage performance production facilities. The paper proceeds as follows. Section 2 briefly recalls the general approach of clustering analysis and Section 3 involves material and methods. Section 4 reports the empirical results and discussions. The section covers the conclusion and policy implications.

Literature Review

Clustering Distance Measures

A distance matrix informs the classification of observations into groups or clusters. The distance measure defines the similarity between two-element (*x, y*) and it greatly influences the shape of the clusters since two objects seem to be not highly correlated due to different characteristics of production. One of the common methods for distance measures is Euclidean distance, which is computed in Equation (1) as:

$$\text{Euclidean distance: } \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

The observations are standardized before clustering to eliminate the distinction of variability on the distance measure. Based on the Euclidean distance, each vector x and y has its mean as $\bar{x}$ and $\bar{y}$ and the distance between x and y is written as (x, y) The aspect of a distance measure is generally linked to the scale or unit based on its measurement. Hence, variables should be standardized before gauging the inter-observation dissimilarities to default distinct scales. The functional relationship can be computed in Equation (2) as:

$$d_{euc}(x,y) = \sqrt{2m[1 - r(x,y)]} \tag{2}$$

where x and y denote two standardized m-vectors under zero mean and unit length.

For the multivariate data analysis of the matrix-like objects, data are transformed using the so-called mean-centering in which they are standardized with zero means and unit variances. Then, the enhanced distance matrix computation and visualization can be carried out. This technique needs criteria to gauge the dissimilarity between the observations.

Cluster Analysis

Cluster analysis is an exploratory technique to explore a system of observations where members inside a group combine specific properties in common. The classified observations into groups are relatively homogeneous within themselves and relatively heterogeneous outside its cluster (Landau & Chis Ster, 2010; Norusis, 2010; Yim & Ramdeen, 2015). There are several types of clustering algorithms.

K-means Algorithm

K-means is a centroid based clustering model. This method operates an iterative clustering algorithm for partitioning a given relevant data into a set of k groups in which k is assigned the proper number of clusters pre-specified by the analyst. It assorts objects in multiple groups, such that these items within the cluster are similar (high intra-class similarity), while other objects outward clusters are dissimilar (low inter-class similarity). Hartigan-Wong algorithm (1979) prescribed the total within-cluster variation in Equation (3) as:

$$W(C_k) = \sum_{x_i \epsilon C_k}(x_i - u_k)^2 \tag{3}$$

where x_i denotes a data point under the cluster C_k and u_k refers to the mean value of the points belonging to the cluster C_k. The individual observation (x_i) is particularized to a given cluster such that the sum of squares (SS) distance of the observation corresponding to their imposed cluster centers (u_k) is minimized. The total within-cluster variation, which measures the concision or goodness of the clustering, is defined in Equation (4) as:

$$tot.withiness = \sum_{k=1}^{k} W(C_k) = \sum_{k=1}^{k} \sum_{x_i \epsilon C_k} (x_i - u_k)^2 \quad (4)$$

Hierarchical Agglomerative Clustering

Hierarchical Agglomerative is a connectivity-based clustering model, which is shaped by connecting data points due to their distance. A dendrogram displays an interpretable visualization of the algorithm and dataset. Ward's Agglomerative Hierarchical Clustering Method seen in Equation (5) was employed. For observations *I, j,* and *k*, these assume:

$$d(i,j) > 0; d(i,j) = 0; \Leftrightarrow i = j; d(i,j) = d(j,i); d(i,j) \le d(i,k) + d(k,j). \quad (5)$$

For an observation set, *I,* associated with $, i, j, k \in I$, the distance can be described as a mapping from the Cartesian product of the dataset into the positive reals: $d: I \times I \longrightarrow \mathbb{R}^{+}$. A dissimilarity generally represents a distance without the triangular inequality qualification $(d(i,j) \le d(i,k) + d(k,j), \forall i,j,k)$. Lance and Williams (1967) utilized the term "an *(i, j)*- measure" for a dissimilarity, given i stands for observation vectors and q reveals the cluster and $q*$ refers to the cluster's center. According to observation i in a cluster q and a distance *d,* we determine a mass associated with observation *i, p(i).* Ordinarily, we assign $p(i) = \frac{1}{|q|}$ when $i \in q$ over cluster cardinality of the cogent cluster and the cluster's center is written as $q^* = \frac{1}{|q| \sum_{i \in q} i}$. The Euclidean distance squared applies norm $\|\cdot\|: if\ i, i' \in \mathbb{R}^{|J|}$, this dataset provides value on attributes $j \in \{1,2, \ldots, |J|\}$, J denotes the attributed set, $|\cdot|$ is cardinality, then $d^2(i,i') = \|i - i'\|^2 = \sum_j (i_j - i'_j)^2$ (Murtage & Legendre, 2014). According to the agglomerative clustering Ward's scheme, the input points are given as vector in Euclidean space with the Euclidean distance as a dissimilarity measure (Müllner, 2013). The distance update formula for $d(i \cup j, k)$ is defined in Equation (6) as follows:

$$\sqrt{\frac{(n_I + n_k)d(I,K)^2 + (n_j + n_k)d(J,K)^2 - n_K d(I,J)^2}{n_I + n_J + n_K}} \quad (6)$$

and the cluster dissimilarity between clusters A and B is expressed in Equation (7) as:

$$\sqrt{\frac{2|A||B|}{|A|+|A|}} \cdot \cdot \left\| \vec{C}_A - \vec{C}_B \right\|_2 \quad (7)$$

where i, j are two clusters joined into a new cluster, and *k* is another. Let n_I, n_J, n_K denote the sizes of cluster *i, j, k*, respectively. The notations $\vec{C}_A$, $\vec{C}_B$ point to the centroid of cluster *A* or *B*.

Model-Based Clustering

Model-based clustering assumes that the data relies on formal models. Sample observations emerge from a distribution involving a mixture of many components (Fraley, Adrian, Brendan, & Luca, 2012). For n-dimensional observations y, the likelihood for a mixture model with G components is written in Equation (8) as:

$$\mathcal{L}_{MIX}(\theta_1, \ldots, \theta_G; \tau_1, \ldots, \tau_G | y) = \prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k f_k(y_i|\theta_k), \quad (8)$$

where G denotes the number of clusters, τ_k is the probability of an object being in the k^{th} group, and f_k stands for the density of the k^{th} cluster within parameters θ_k. For Latent class model and EM algorithm, the general form is: $\prod_{k-1}^{G} f_k(y_i|\theta_k)^{z_{ik}}$. The completed data covering the unobserved latent z has the likelihood as: $L_c(y_i, z_i|\theta)$=$\prod_{i=1}^{n} f(y_i, z_i|\theta)$ and the observed data likelihood is as: $L_o(y_i|\theta)$=$\int L(y_i, z_i|\theta)dz$. The conditional expectation of log L_C given the observed data and the current parameter estimation is computed by the E step, which is formulated as:$\frac{\hat{\tau}_k f_k(y_i|\hat{\theta}_k)}{\sum_{j=1}^{G} \hat{\tau}_j f_j(y_i|\hat{\theta}_j)}$. The expectation is maximized after putting in the expectation of Z_{ik} using the *M* steps in Equations (9) and (10) as follows:

$$\prod_{k=1}^{G} f_k(y_i|\hat{\theta}_k)^{z_{ik}}, \quad (9)$$

$$l(\theta_k, \tau_k, z_{ik}|X) = \sum_{i=1}^{n} \sum_{k=1}^{G} z_{ik} \log[\tau_k f_k(y_i|\theta_i)]. \quad (10)$$

Methodology

Data Analysis

National Economic and Social Development Board (NESD, 2013) reported the national accounts of Gross Provincial Product (GPP) (11 sectors). Due to Songkhla being designated for SEZ development, we inquired about production clustering analysis. The determinants of three macroeconomic parameters, in which we explore the production clusters to facilitate the enhancement of competitiveness, were employed as a proxy:

1. The average annual growth rate (AAGR)- is calculated in chain volume measures in Million Baht and obtained from NESD over 10 time periods since this can inform long-term trends. The AAGR, which is calculated by taking the arithmetic mean of a series of growth rates, is truly helpful for analyzing trends in terms of the 11 production sectors measured to GPP. Since SEZ development plans should be determined as long-term circumstances, the AAGR then can practically indicate industrial activities across multiple periods. Moreover, it can convey to policymakers or investors an idea of which direction of economic performance is likely for that particular measure;

2. Share of the production sector- in percent of GPP and obtained from NESD. The Share of the production sector can be employed to appraise the effectiveness compared to other production sectors. This proxy imparts the scales of each production sector in contribution to economic growth and allows details of quantifying the impact strategies or tactical execution on economic achievement;

3. Employment rates- by the means of the employment rate at Songkhla and acquired via the National Statistical Office. The employment rates refer to a measure of the extent to which available labor resources are being employed. It is normally recognized as a key indicator of the performance of a country's labor market. If labor resources are employed, individuals and households earn wages, and ultimately the nation achieves economic development.

Consequently, the three macroeconomic parameters, which are determined as a part of the influencing in terms of the institutional change in the field of economic policies, were chosen.

Data Collection

These relevant datasets are calculated as the central values from 2008 to 2018, totally eleven years, to be received for the annual growth rate of ten years. Before these periods, most of the statistical datasets had not been fully recorded. The standardization of variables was carried out when using Euclidean distance as the dissimilarity metric. Since the proper clusters to carry out are unknown, the application of clustering analysis starts at first to determine optimal number of clusters, then the clustering model is rerun with a certain group (Burns and Robert, 2008; Chávez, Torres, & Torres, 2016; Lipták, Klasová, & Kováč, 2015).

For application software, the R program for statistical computation compiles a wide variety of cluster analyses. In this article, the K-Means clustering algorithm is computed using the usage of k-means (x = numeric matrix of data, centers = a set of an initial cluster, iter.max = the maximum number of permitted iterations, nstart = proper random sets, algorithm = c (either “Hartigan-Wong” “Lloyd”, “Forgy”, or “MacQueen”), trace = FALSE). For Ward Hierarchical clustering, this function reckons and converts the distance matrix associated with the specified distance measure to estimate the distances between the rows of a data matrix using the usage of the dist (x = a numeric matrix, method = either “euclidean”, “maximum”, “manhattan”, “canberra”, “binary” or “minkowski”, diag = logical number implying the diagonal of the distance matrix, upper = logical number implying the upper triangle of the distance matrix, p = the power of the Minkowski distance). For Model-Based Clustering, this function calculates Gaussian finite mixture models fitted via the EM algorithm using the usage of the mclust (x = a numeric matrix). It reports Bayesian Information Criterion (BIC), classification, uncertainty and density information.

Results

Clustering Distance Measurements

At the beginning of the cluster analysis process, the standardization method was executed through the dataset to be transformed as *zero mean* and unit *variance*. Table 1 shows the distance measures on both the full dataset of 11 production sectors using the Euclidean method and those standardized data are comparative.

The visualization of the distance matrices in which the color shade indicates the degree of dissimilarity between variables- the pure blue (value = 0) implies the strong similarity, $dist(x_i, y_j) = 0$; and the pure red (value = 3) implies the weak similarity. It can be seen that the HOTE sector appears with outstandingly large similarities with the TRAN, ELEC, FINAN, and ESTE. On the other hand, there are fairly similar results in terms of AGRI, MIN, MANUM, FISH, and CONS as depicted in Figure 1. Moreover, based on the visualized distance matrices, the datasets were classified into four or five subgroups.

Identify Optimal Clusters

Within a simple criterion to identify the pre-specified five classes as centroids, the K-means clustering approach was conducted. The cluster experiment and the centroid's updated array are iteratively to reach convergence. (Figure 2A) displays the total within-cluster sum of square (WSS) as the sum of squared distanced Euclidean distances. It can be noted that the initial values of K (K = 1 to K = 3), adding a

Table 1 Distance matrices among the 11 production sectors

	AGRI	FISH	MIN	MANU	ELEC	CONS	SALE	HOTE	TRAN	FINAN
FISH	2.483									
MIN	3.632	3.213								
MANU	3.017	3.516	1.545							
ELEC	3.648	2.725	2.324	3.443						
CONS	2.464	1.205	2.549	3.1282	1.586					
SALE	1.432	2.822	2.523	1.982	2.886	2.262				
HOTE	3.394	3.371	2.501	3.261	1.227	2.177	2.384			
TRAN	2.975	2.094	1.825	2.753	0.861	0.995	2.247	1.498		
FINAN	3.719	2.932	1.982	3.172	0.441	1.804	2.803	1.145	0.891	
ESTE	3.523	2.673	1.828	2.992	0.496	1.573	2.647	1.275	0.624	0.289

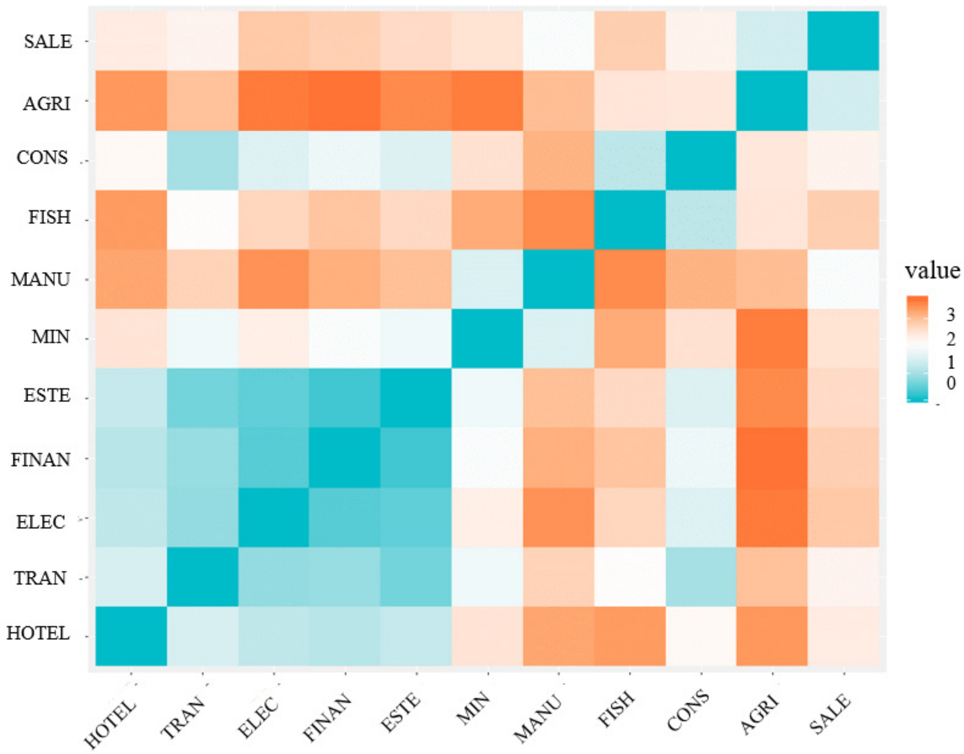

Figure 1 Visualization for the distance matrices among 11 production sectors

group, diminish sharply the WSS criterion. The improvement is low at K = 4 and seems to be the right subgroups.

To verify the optimal partitions, we applied the NbClust criteria, which conducts 30 indices of the distance measures such as Euclidean, Maximum, Manhattan, Canberra, Binary or Minkowski to compute the dissimilarity matrix (Charrad, Ghazzali, Boiteau, Niknafs, & Niknafs, 2015). According to Figures 2B) and 2C), the significant knee in Dindex and the significant peak of the second difference Dindex are meaningful at K = 4. From frequency among all indices (i.g. 9 proposed 4; 5 proposed 2; 4 proposed 6 etc.), the majority principles recommended the best clustering scheme at four. Furthermore, the Elbow method and the Gap statistic accompanied by the bootstrap replications of 500 also approve for the partition at K = 4; whereas the Silhouette method suggests the possible segmentation at K = 3 as seen in Figures 2D) and 2F). Consequently, the appropriate partition at K = 4 was chosen since this is confirmed by those complementary tests.

Fundamentally, the total intra-cluster variation or the total within-cluster sum of square (WSS), gauges the compactness of the clustering. These imply the lower the measured distances, the better the cluster performance. To consider intuitions behind partitioning, a bend of a curve informs a heuristic optimization to select a point, where diminishing returns are no longer worth the additional cost. At the cutoff point, adding another cluster does not create much better modeling of the data so that the knee of a curve (belonging to the WSS, Index in Nbclust, and Elbow) is computationally and automatically chosen based on the optimal point accompanied by trials and errors via the R packages. For the second difference Dindex, the location of the maximum is determined as the appropriate number of clusters (K = 4). For the Gap statistic, it starts to slightly drop after five clusters, introducing that there are four well-separated clusters. Consequently, the partition at K = 4 seems to be the initial cluster and this is consistent with those complementary tests. Henceforward, the preliminary cluster (K = 4) is chosen.

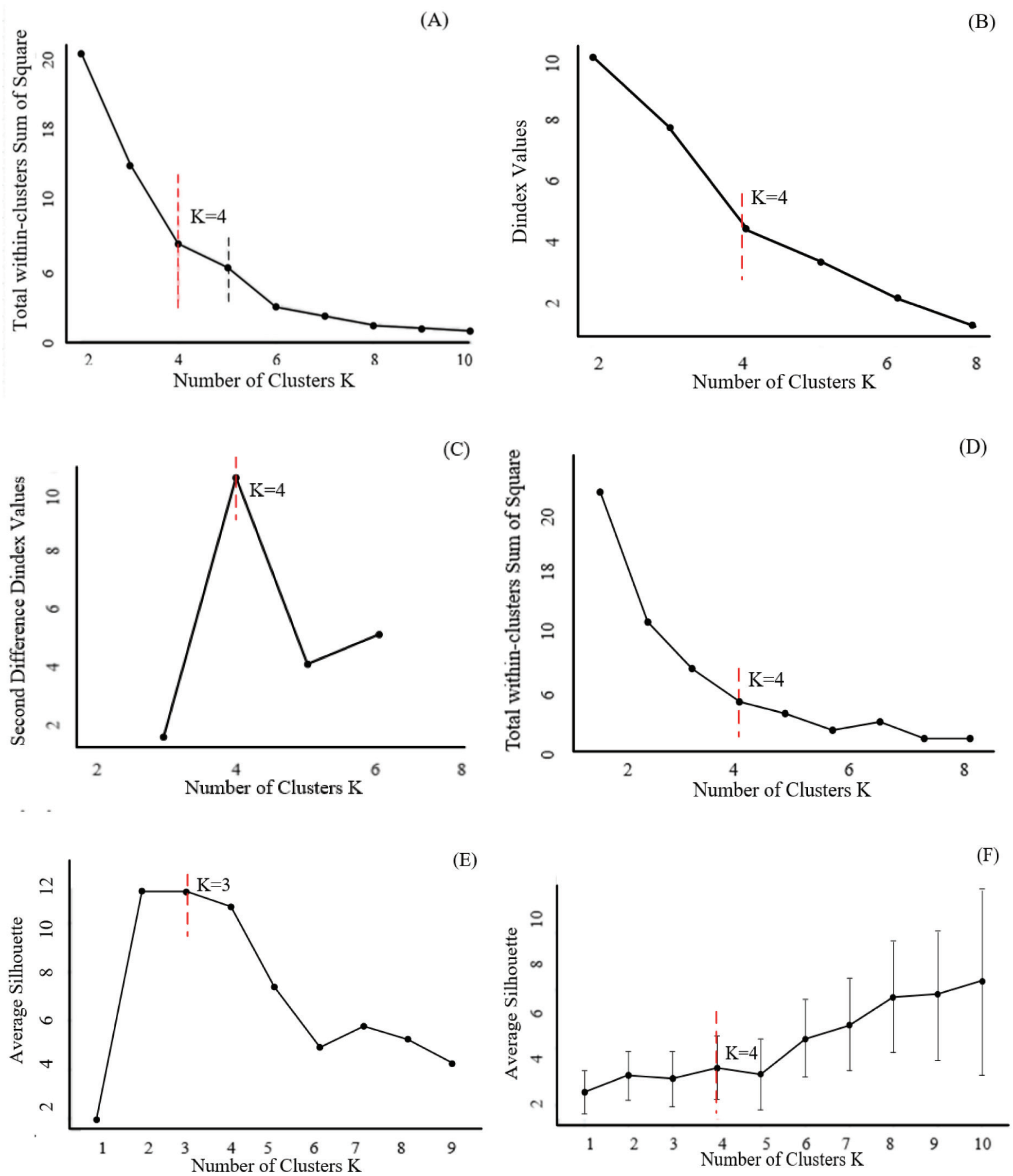

Figure 2 Determination the best number of clusters; (A) WSS value, (B) Index in NbClust, (C) Secound Different Dindex NdClust (D) Elbow Method (E) Silhouette Method and (F) Gap Statistic

Cluster Solutions

K-means algorithm

The K-Means Clustering with 4 partitions was carried out as demonstrated in Figure 3. The first cluster consisted of AGRI and SALE. The second cluster was made up of ELEC, HOTE, TRAN, FINAN, and ESTE. The third cluster was FISH and CONS. Finally, the fourth cluster comprised of MANU and MIN, which are the large scale of production volumes in Songkhla's SEZ. Table 2 presents the results of the K-Means Clustering with 4 partitions and the WSS segregated by the clusters 1 to 4 are 1.026, 1.828, 0.726, and 1.194, respectively. Moreover, the Between Sum of Squares (BSS) to Total Sum of Squares (TSS) or the BSS/TSS ratio is at 84.1 percent, indicating a goodness of fit for the classification K-Means approach since the proposed cluster has precisely the properties of internal cohesion against external decomposition.

Hierarchical agglomerative

The uncertainty in hierarchical cluster demonstration was successively assessed using the Ward method. These findings are consistent with the K-mean clustering that the AGRI and SALE are merged and the MANU and the MIN separately stand alone in its hedge. The FISH and CONS are in the third group. The ELEC, HOTE, TRAN, FINAN, and ESTE belong to the last subgroup as presented in Figure 4A. Furthermore, we conducted the Hierarchical clustering algorithm accompanied by the Multiscale Bootstrapping 1,000 duplications using the Ward method and Euclidean-based dissimilarity matrix as shown in Figure 4B. The two types of bootstrap probability (BP), which coincides with the frequency of identified bootstrap replication as well as the approximately unbiased (AU) probability values (p-values) by multiscale bootstrap repeating, were computed. Within a cluster with AU $p > .90$, the hypothesis that "the cluster does not occupy" is rejected with significance level .10. Therefore, these four highlighted clusters tend to exist in Songkhla's SEZ.

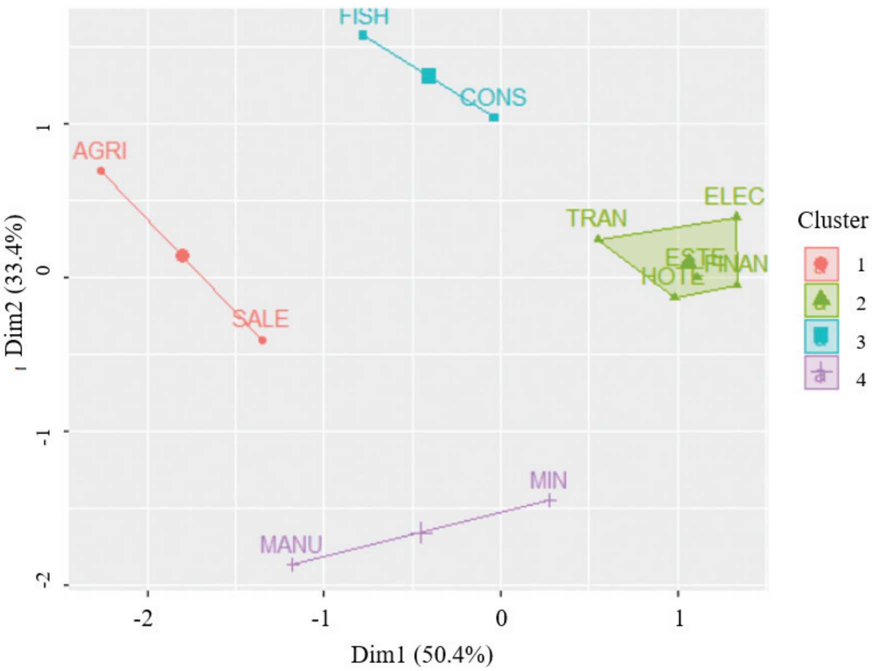

Figure 3 K-Means Clustering with four partitions

Table 2 Results of the K-Means Clustering with 4 partitions

Cluster	Clustering vector	AAGR	Share of production	Employment
1	AGRI, SALE	0.698	0.304	1.806
2	ELEC, HOTE, TRAN, FINAN, ESTE	0.716	0.597	0.578
3	FISH, CONS	1.388	0.625	0.221
4	MANU, MIN	0.296	1.811	0.139

Note: The within-cluster sum of squares (WSS) by cluster: 1.026, 1.828, 0.726, and 1.194
The BSS/TSS ratio = 84.1 percent.

Model-based clustering

In addition, we also applied the Model-Based clustering and classification estimation, which speculates a distribution. Unlike the K-Means and the Hierarchical agglomerative criteria, the Model-Based clustering assigns the observed data as a Gaussian finite mixture accompanied by different covariance structures and different numbers of mixture components within a group via the EM algorithm and the Bayesian Information Criterion (BIC) (Farley et al, 2012). Figure 5A displays the best model with the ellipsoidal distribution, equal volume and shape (EEV; BIC = -1.126) with the optimal clusters of the 5-component mixture. Cluster 1 is AGRI and SALE, Cluster 2 consists of CONS, FINAN, ESTE, HOTE, and FISH. Cluster 3 is MIN and cluster 4 MANU. Cluster 5 includes ELEC and TRAN (Figure 5B).

Clustering validation measure

To measure the clustering validation, the internal evaluation schemes including average Silhouette width, and Dunn index were carried out as summarized in Table 3. The average Silhouette width is 0.73 (near to 1), indicating the well-clustered observations. Moreover, the higher Dunn index indicates the better clustering, determining the four clusters at 0.983 against the five clusters at 0.232. Therefore, the clustering analysis with 4 partitions based on the clustering validation indices seems to be satisfied and is the appropriate solution since this criterion suits every outlier in its cluster.

We concisely identify each segment. For Cluster 1 (AGRI and SALE), the AGRI is one of the important engines for economic growth, accounting approximately 18 percent of GPP. About 44 percent of labor forces are engaged in the agriculture sector. The proportion of the SALE contributes around 11 percent of GPP. As the sequence, if the agriculture employers earn more revenue, this can simultaneously stimulate more expenditure on the SALE. For Cluster 4 (MANU and MIN), the MANU

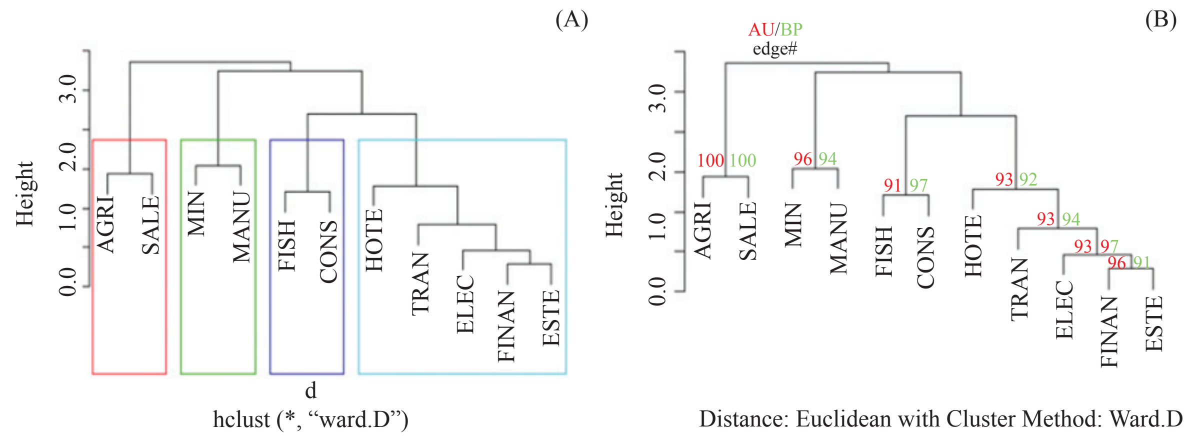

Figure 4 Cluster Dendrograms with four partitions; (A) Cluster Dendogram, (B) Cluster Dendogram with AU/BP

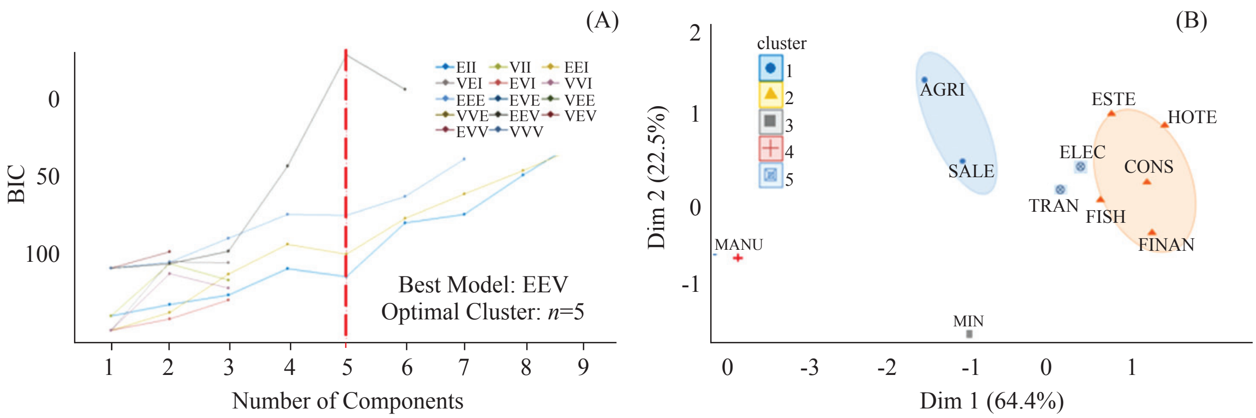

Figure 5 Model-Based clustering with the five optimal clusters; (A) Model Selection, (B) Classification

Table 3 Results of clustering validation measurement

Clustering validation indices	Values	Interpretations
Average Silhouette width	0.73 [0.64200, 0.482, 0.547, and 0.731]	A strong structure
The Dunn index: 4 clusters Vs. 5 clusters	0.983 Vs. 0.232	A higher Dunn index for 4 cluster
average-between cluster	3.418	as large as possible
average-within cluster	0.046	as small as possible

becomes the main driver of economic growth, accounting for more than 22 percent of GPP. The proportion of the Min exhibits about 13 percent of GPP. The MANU relates to agro-based industries such as rubber, palm oil or, sea-food processing. These industries employ agricultural products as their raw materials in which the AGRI (from the Cluster 1) has strong potential to deliver bulky raw materials and supplies. Moreover, the MIN involves petrochemicals processing. Cluster 4 should be supported to be industrial development by setting up agriculture production zones to offer raw material for the agro-industry and the petrochemicals processing. Moreover, this Cluster should serve as one of the nine industrial centers since the Southern Seaboard development plan can promote the SEZ as a long-term economic base within Thailand. Cluster 2 comprises of the ELEC, HOTE, TRAN, FINAN, ESTE. These five sectors are regarded as the service segment that encourages private investors to invest in tourism activities and real estate investment. Cluster 3 consists of the FISH and CONS which take a trivial share of GPP.

Conclusion and Recommendation

For specifying the optimal clusters, the majority rules based upon the NbClust method suggest the four subgroups identified, and it is supported by the Elbow method and the Gap statistic. According to the K-means Algorithm and the Hierarchical clustering algorithm, Cluster 1 is AGRI and SALE, the largest Cluster 2 consists of ELEC, HOTE, TRAN, FINAN, ESTE, and Cluster 3 is FISH and CONS. Cluster 4 is MANU and MIN. Unlike those two algorithms, the Model-Based clustering informs the different scheme of five clusters, comprising of Cluster 1 (AGRI and SALE), Cluster 2 (CONS, FINAN, ESTE, HOTE, FISH), Cluster 3 (MIN), Cluster 4 (MANU), and Cluster 5 (ELEC and TRAN). Moreover, the clustering validation measurements of the Dunn index and Silhouette width were adopted to approve the appropriate clustering analysis with four segmentations. All tier clusters, particularly in Cluster 1 (AGRI and SALE) and Cluster 4 (MANU and MIN), precisely reflect recent economic performance at Songkhla's SEZ.

The key role of the clustering approach of the SEZ is to ensure optimal production and to increase the sustainability of economic development since clusters based on regional economies can integrate strong internal dynamics within-cluster and can verify interconnection within a specific field production. Regions with strong clusters and solid connections impart innovative leaders. Suitable regional production clusters generate several advantages due to higher productivity and competition. Consequently, to gain advantages of the SEZ development, policy-makers can consider the recommended production clustering in Songkhla's SEZ, especially for the two strong clusters (Cluster 1- AGRI and SALE and Cluster 4- MANU and MIN), and the others supportive clusters can enhance the level of competitiveness in a regional production and a global market instead of isolated production sectors (Türkcan, Tunalı, & Kaya, 2009). Furthermore, proper clusters push regional and national economy through direct and indirect influences such as research and development (R&D) in business or patent output.

Conflict of Interest

There is no conflict of interest.

References

Becattini, G., Bellandi, M., Dei Ottati, G., & Sforzi, F. (2003). *From industrial districts to local development: An itinerary research*. Cheltenham, UK: Edward Elgar Publishing Limited.

Charrad, M., Ghazzali, N., Boiteau, V. Niknafs, A., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, *61*(6), 1–36. doi: 10.18637/jss.v061.i06

Chávez, J. C., Torres, A. I., & Torres, M. C. (2016). Hierarchical cluster analysis of tourism for Mexico and the Asia-Pacific Economic Cooperation (APEC) countries. *Torismo EM Análise*, *27*(2), 235–255. doi: 10.11606/issn.1984-4867.v27i2p235-255

Cherdchom, P., Prommin, P., & Romyen, A. (2016). *Reports of research on growth of small and meduim enterprises under special economic development zone: Case study Sadao district Songkhla province*. Songkhla, Thailand: Faculty of Economics, Prince of Songkla University. [in Thai]

Cossentino, F., Pyke, F., & Sengenberger, W. (1996). *Local and regional responses to global pressure: The case of Italy and its industrial districts*. Geneva, Switzerland: International Institute for Labour Studies.

Fraley, C., Adrian E. R., Brendan M. T., & Luca, S. (2012). *Mclust version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation*. Retrieved from https://stat.uw.edu/sites/default/files/files/reports/2012/tr597.pdf

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, *28*(1), 100–108. doi: 10.2307/2346830

Lance, G. N., & Williams, W. T. (1967). A general theory of classificatory sorting strategies: 1. Hierarchical systems. *Computer Journal*, *9*(4), 373–380. doi: 10.1093/COMJNL/9.4.373

Landau, S., & Chis Ster, I. (2010). Cluster analysis: Overview. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (3rd ed., pp. 72–83). Oxford, UK: Elsevier Ltd.

Lipták, F., Klasová, S., & Kováč, V. (2015). Special economic zone constitution according to cluster analysis. *Procedia Economics and Finance*, *27*, 186–193. doi: 10.1016/S2212-5671(15)00988-0

Müllner, D. 2013. Fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. *Journal of Statistical Software*, *53*(9), 1–18. doi: 10.18637/JSS.V053.I09

Murtage, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *Journal of Classification*, *31*, 274–295. doi: 10.1007/s0035-014-9161-z

National Economic and Social Development Board [NESD]. (2013). *Gross regional and provincial product: Chain volume measures 2013 Edition*. Retrieved from https://www.nesdc.go.th/nesdb_en/ewt_w3c/ewt_dl_link.php?filename=national_account&nid=4317

Norusis, M. J. (Ed.). (2010). Cluster analysis. In *PASW statistics 18 statistical procedures companion* (pp. 361–391). Upper Saddle River, NJ: Prentice Hall. Retrieved from http://norusis.com/pdf/SPC_v13.pdf

Porter, M. P. (2000). Location, competition and economic development: Local clusters in a global economy. *Economic Development Quarterly*, *14*(1), 15–24. doi: 10.1177/089124240001400105

Propris, L. D., & Driffield, N. (2005). The importance of clusters for spillovers from foreign direct investment and technology sourcing. *Cambridge Journal of Economics*, *30*(2), 277–291. doi: 10.1093/cje/bei059

Romyen, A., Liu, J., Sriboonchitta, S., Cherdchom, P., & Prommee, P. (2019). Assessing regional economic performance in the Southern Thailand special economic zone using a Vine-COPAR model. *Economies*, *7*(2), 30. doi: 10.3390/economies7020030

Türkcan, B., Tunalı, E., & Kaya, A. A. (2009). *Industrial clusters as a regional development tool: A spatial analysis on Turkey*. Paper presented at EconAnadolu 2009: Anadolu International Conference in Economics, Eskişehir, Turkey. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.521.9793&rep=rep1&type=pdf

World Bank. (2009). *World development report 2009*. Washington, DC: Author.

Yim, O., & Ramdeen, K. T. (2015). Hierarchical cluster analysis: Comparison of three linkage measures and application to psychological data. *The Quantitative Methods for Psychology*, *11*(1), 8–21. doi: 10.20982/tqmp.11.1.p008