Jurnal Sistem Teknik Industri
Journal homepage: https://talenta.usu.ac.id/jsti

Unlocking Indonesia's Maritime Potential: Optimizing Hub Port Development using a Principal Component Analysis and K-Means Clustering

Dwi Adi Purnama[1*], Putri Citra Marifa[2], Riadho Clara Shinta[3]
[1] *Departement of Industrial Engineering, Universitas Islam Indonesia, Yogyakarta, 55584, Indonesia*
[2] *PT PLN (PERSERO), Jakarta, 12160, Indonesia*
[3] *Otoritas Jasa Keuangan (OJK), Jakarta, 10710, Indonesia*
[*]Corresponding Author: dwiadipurnama@uii.ac.id

ARTICLE INFO

Article history:
Received 19 October 2024
Revised 25 December 2024
Accepted 5 January 2025
Available online 29 January 2025

E-ISSN: 2527-9408
P-ISSN: 1411-5247

How to cite:
Purnama, D. A., Marifa. P. C., & Shinta. R. C., (2025). Unlocking Indonesia's Maritime Potential: Optimizing Hub Port Development using a Principal Component Analysis and K-Means Clustering. *Jurnal Sistem Teknik Industri*, 27 (1), 35-46.

http://doi.org/10.32734/register.v27i1.idarticle

ABSTRACT

Indonesia as a maritime country, most of the transportation in Indonesia is carried out by using sea transportation around 88%. However, the Logistics Performance Index (LPI) of Indonesia is alarming and declining from 3.15 in 2018 to 3.0 in 2023, particularly in Timeliness (3.7 to 3.0) and Tracking & Tracing (3.3 to 3.0). Therefore, it needs a comprehensive improvement strategy, especially to optimize the hub port connectivity. This study would like to establish the new hub port connectivity in Indonesia to increase the effectiveness and efficiency of the port. This study use Principle Component Analysis (PCA) to determine the variable, eliminate the correlation among variables to obtain the new variables as the clustering determinants factor, and use Clustering K-Means to group similar characteristics of the alternative port that can be used as the main port. The result of Principle Component Analysis (PCA) which has gained new variables are regional spatial system, national transportation system, national defense, operational cost, and port services. The result of K-Means Clustering, Tanjung Priok, Panjang, Tanjung Emas, Tanjung Perak, Palembang, Teluk Bayur, and Belawan are the main port or hub port connectivity to increase the effectiveness and efficiency in the sea transportation. Tanjung Priok and Panjang can connect Java and Sumatra, while Tanjung Emas and Tanjung Perak can link Java and South Kalimantan. Palembang, Teluk Bayur, and Belawan can also play important roles in their respective regions. By investing in infrastructure and improving connectivity, Indonesia can enhance the efficiency and competitiveness of its maritime trade.

Keyword: Optimize Hub Port Development, Principal Component Analysis, K-Means Clustering, Hub Port Connectivity, Effectiveness and Efficiency

ABSTRAK

Indonesia sebagai negara maritim, sebanyak 88% transportasi di Indonesia dilakukan dengan menggunakan transportasi laut. Namun, Indeks Kinerja Logistik (LPI) Indonesia masih mengkhawatirkan dan terus menurun dari 3,15 pada tahun 2018 menjadi 3,0 pada tahun 2023, terutama pada aspek Timeliness (3,7 menjadi 3,0) dan Tracking & Tracing (3,3 menjadi 3,0). Oleh karena itu, diperlukan strategi perbaikan yang komprehensif, terutama untuk mengoptimalkan konektivitas pelabuhan hub. Penelitian ini ingin membangun konektivitas pelabuhan hub baru di Indonesia untuk meningkatkan efektivitas dan efisiensi pelabuhan. Penelitian ini menggunakan Principle Component Analysis (PCA) untuk menentukan variabel, menghilangkan korelasi antar variabel untuk mendapatkan variabel baru sebagai faktor penentu clustering, dan menggunakan Clustering K-Means untuk mengelompokkan karakteristik yang sama dari pelabuhan alternatif yang dapat digunakan sebagai pelabuhan utama. Hasil Principle Component Analysis (PCA) yang diperoleh 5 variabel adalah *regional spatial system*, *national transportation system*, *national defense*, *operational cost*, dan *port services*. Hasil Clustering K-Means, Tanjung Priok, Panjang, Tanjung Emas, Tanjung Perak, Palembang, Teluk Bayur, dan Belawan merupakan konektivitas pelabuhan utama atau hub port untuk meningkatkan efektivitas dan efisiensi transportasi laut. Tanjung Priok dan Panjang dapat menghubungkan Pulau Jawa dan Sumatera, sedangkan Tanjung Emas dan Tanjung Perak dapat menghubungkan Pulau Jawa dan Kalimantan Selatan. Palembang, Teluk Bayur, dan Belawan juga dapat memainkan peran penting di daerahnya masing-masing. Dengan berinvestasi di bidang infrastruktur dan meningkatkan konektivitas, Indonesia dapat meningkatkan efisiensi dan daya saing perdagangan maritimnya.

Keyword: Optimalisasi Pengembangan Hub Pelabuhan, Principal Component Analysis, K-Means Clustering, Konektivitas Hub Pelabuhan, Efektivitas dan Efisiensi

1. Introduction

Indonesia, known for its diverse culture and maritime heritage, possesses abundant natural resources and significant maritime potential. Marine resources and marine transportation are key sectors where Indonesia can leverage its advantages. Sea transportation is the dominant mode of transportation in Indonesia, accounting for approximately 88% of all transportation activities [1]. Its efficiency and payload capacity make it a preferred choice for carrying goods. To fully realize the potential of its maritime sector, Indonesia must focus on improving its marine transportation policies and management. This includes measures such as container maintenance and enhancing overall logistical performance. By optimizing these aspects, Indonesia can aim to reduce the cost of national logistics, leading to greater competitiveness and economic growth.

Indonesia's existing logistics performance is still alarming, necessitating the development of a comprehensive improvement strategy. According to the Logistics Performance Index (LPI), Indonesia falls into the partial performance group year after year, with LPI dropping from 3.15 in 2018 to 3.0 in 2023 [2]. This also indicates a drop in worldwide position from 43 to 63, with Singapore, Malaysia, Thailand, the Philippines, and Vietnam staying ahead of it in terms of ASEAN countries. Indonesia's Logistics Performance Index (LPI) has declined in recent years, particularly in the areas of Timeliness (from 3.7 to 3.0) and Tracking & Tracing (from 3.3 to 3.0). The decrease in Indonesia's punctuality score is likely to be caused by bottlenecks at the Port due to supply chain interruptions that occurred subsequently.

Ports are essential for facilitating trade and economic development in Indonesia. While strategic port development has been a priority, the high cost of logistics in eastern Indonesia compared to the west remains a concern [1]. Analysts suggest that limited port connectivity is a major contributing factor. By improving port connectivity through infrastructure upgrades and policy initiatives, Indonesia can significantly reduce logistics costs and enhance its overall economic competitiveness. Moreover, container shipping is crucial for global trade, but it faces various risks, including pandemics, conflicts, terrorism, and natural disasters [3], [4], [5], [6], [7], [8], [9]. To strengthen Indonesia's container shipping sector, it's essential to diversify trade routes, improve port security, develop emergency plans, invest in infrastructure, and collaborate with neighbouring countries. These measures can help mitigate the impact of disruptions and ensure a more resilient and sustainable supply chain.

In the other side, despite a decline in the average gross tonnage of cargo shipped by sea in 2022 to 1851,23 million gross tonnes of cargos (decreased by 9.38%) [10], the volume of cargo loaded on domestic voyages at 25 strategic ports in Indonesia experienced a 4.20% increase in 2022 compared to the previous year. This growth was particularly notable at the three main ports: Belawan, Tanjung Priok, and Makassar. Belawan saw a significant surge of 595.71%, while Tanjung Priok and Makassar recorded increases of 9.78% and 20.10%, respectively. Several strategic ports in Indonesia experienced significant increases in cargo volume in 2022. Teluk Bayur, Dumai, Panjang, Batam, Tanjung Emas, Tenau Kupang, Pontianak, Banjarmasin, Jayapura, and Biak all recorded notable growth rates. For example, Tanjung Emas saw a substantial increase of 494.87%, while Jayapura experienced a more modest growth of 2.79%. However, not all ports experienced growth. Tanjung Perak Port saw a decrease of 22.11%, and several other ports, including Lhokseumawe, Palembang, Tanjung Pinang, Banten, Benoa, Balikpapan, Samarinda, Bitung, Ambon, and Sorong, also recorded declines in cargo volume. Pekanbaru Port had no loading activity during the period.

Therefore, based on the potential dan the problem in Indonesian port including the performance logistic in Indonesia, there need to optimize the hub port connectivity to increase effectiveness and efficiency, and to minimize the transportation cost and transportation time while shipping the container to the distributor. According to Low et al. [11] the researcher decided to minimize the transportation cost and time by implement a concept to build up a main hub for another ship transit through largest hub port in Western Indonesia which support the effectiveness and efficiency of logistic system in Indonesia.

This paper proposed to optimizing hub port development using a Principal Component Analysis (PCA) and K-Means Clustering to increase effectiveness and efficiency and minimize the transportation cost and transportation time while shipping the container. Previous research on port connectivity has only focused on identifying the density of connections between ports in eastern Indonesia and analyzing freight and cargo connectivity at airports. Studies by Zaman et al.[1], Kim and Park [12], Manzano et al. [13], Bai et al. [14], Asadabadi & Miller-Hooks [15], Guo et al. [16], He et al. [17] and Kang et al. [18] provide valuable insights into the factors influencing port and airport connectivity, which can be applied to inform strategies for improving Indonesia's port infrastructure and connectivity. In the other side, The majority of studies previously also related to container maritime transportation systems [19], that typically used AIS data and service schedules from various shipping companies to examined the network's structure [20], the hub port's mining [21] and community structure [22], comprehensive assessments on the vulnerability and resilience of the Global Container Shipping Network [23]. As a result, in the best of our knowledge, this paper is the first study to optimizing hub port connectivity using PCA and Clustering method and the main contribution is to determine the determine the alternatives port that can be used as the main port or hub port to increase the effectiveness and efficiency from the existence port.

2. Method

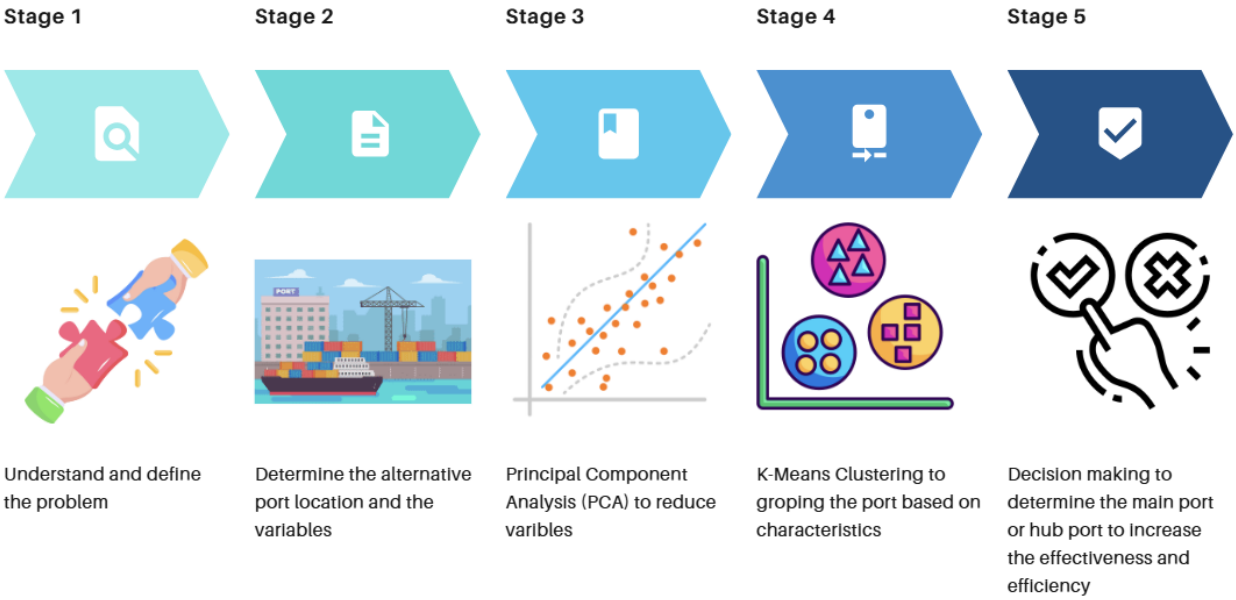

Figure 1. Flowchart Research

Research methodology which used in this research is a case study from data of crowded port in Indonesia. By collecting the marine traffic data, it can show the problem of Indonesian port. Thus, this paper purpose is to determine the main port or hub connectivity using Principal Component Analysis (PCA) and Clustering also to increase the effectiveness and efficiency of the port connectivity in Indonesia. This research through several stages (Figure 1), such as:

Stage 1: In this stage, the researcher did a literature review about the problem which exist in Indonesian port.

Stage 2: In the second stages, the researcher uses the literature review to determine the alternative of port location and determine the variables.

Stage 3: In this third stages, there are data collecting for several port in Indonesia. In this stage also applied principal component analysis (PCA) to reduce the variable [24], eliminate the correlation among variables to obtain the new variables as the clustering determinants factor.

Stage 4 : There will be a variable computation which had been analyzed using principle component analysis for doing the clustering process, until it can be grouped for the ports based on the characteristics [25].

Stage 5 : In the last stages, there should determine the alternatives port that can be used as the main port or hub port to increase the effectiveness and efficiency from the existence port.

For determine the port as the main hub for the ship transit which recognized by Indonesia Government Regulation, there explained the criteria of port location established which shown in the Table 1 based on Sutomo and Soemardjito [26].

Table 1. The location criteria for hub and sub-hub port

Feature	Hub Ports	Sub-Hub Ports
Location	Near international markets and shipping routes	Strategically located within a country
Capacity	Larger capacity to handle larger vessels	Smaller capacity compared to hub ports
Function	Serve as transshipment points for international trade	Support regional development and domestic shipping

According to Table 1, it was explained the location criteria for hub port in Indonesia, then the researcher adjusts the secondary data based on the criteria of hub ports (international ports) to do a calculation.

Clustering and Principal Component Analysis (PCA) are tools for addressing the research topic. Clustering is a method of grouping similar data items. A clustering method divides a large number of data points into a smaller number of groups so that data points from the same group share. Clustering has numerous uses, such as family creation for group technology, image segmentation, information retrieval, web page grouping, market segmentation, and scientific and engineering analysis [27]. Despite this, the statistical technique of unsupervised dimension reduction is commonly utilized in PCA. K-means clustering is a popular data clustering technique for performing unsupervised learning tasks [28].

In this research, there are several variables used during determine the new hub port. The variables and data used are obtained from the secondary data which contained in the journal of Assessment Model of the Port Effectiveness and Efficiency (Case Study: Western Indonesia Region) by [26]. According to Sutomo dan Soemardjito [26], there are seven variables that can be used to assess the effectiveness and efficiency of a port. The variables consist of:

1. Regional Spatial System (RSS) is the network of infrastructure and services within a region that enables the movement of goods and people. A well-integrated RSS ensures the port is connected to its hinterland, facilitating smooth cargo flow and supporting regional economic growth. Moreover, an efficient RSS minimizes transportation costs and time for cargo moving to and from the port, leading to improved overall logistics performance.

2. National Transportation System (NTS) is the interconnected network of roads, railways, waterways, and airways that facilitate the movement of goods and people across a nation. A robust NTS ensures the port is seamlessly integrated into national and international transport networks, enabling efficient cargo movement across the country and beyond. Then, A well-functioning NTS minimizes delays and bottlenecks in cargo transportation, leading to faster turnaround times for ships and reduced overall logistics costs.

3. National Defense (ND) is the military and civilian capabilities of a nation to protect its interests. A strategically located port within the NTS can play a crucial role in national defense by facilitating the movement of troops, equipment, and supplies. Moreover, a well-maintained port infrastructure can quickly mobilize defense assets in times of need, ensuring national security.

4. Operational Cost (OC) is the expenses incurred by a port in its day-to-day operations, including labor, energy, maintenance, and administration. Lower OC indicates efficient port operations, attracting more shipping lines and cargo, and enhancing the port's competitiveness. Moreover, minimizing OC through optimized processes, technology, and infrastructure leads to lower costs for port users, boosting economic activity.

5. Commodity (C) is types of goods handled by a port, such as containers, bulk cargo, or specialized products. A diverse range of commodities handled by the port indicates its ability to serve various market segments and adapt to changing economic conditions. Moreover, specializing in specific commodities can lead to optimized port operations and infrastructure, improving efficiency and reducing costs.

6. Port Service (PS) is the range of services offered by a port, including cargo handling, storage, customs clearance, and other value-added services. A wide range of high-quality PS attracts shipping lines and cargo owners, enhancing the port's competitiveness. Efficient PS, characterized by quick turnaround times and minimal delays, reduces overall logistics costs and improves customer satisfaction.

All of these data had been through the standardization process which used to synchronize the type of data by compare one of the criteria with the others by using numbers in the range of 0-1. The standardization method uses the following formula:

$$\text{Sbi} = (\text{Score i - Minimum Score}_{i-n}) \;/\; (\text{Maximum Score}_{i-n} - \text{Minimum Score}_{i-n}) \quad (1)$$

$$\text{Sci} = (\text{Score i - Minimum Score}_{i-n}) \;/\; (\text{Maximum Score}_{i-n} - \text{Minimum Score}_{i-n}) \quad (2)$$

3. The Clustering Method

Clustering is a data mining process that works by identifying and grouping data that have similar features to the original data and the other data retrieved. The characteristics of data mining techniques are designed to be unsupervised or without direction, meaning they can be used without data training or a teacher and do not require a desired output [32].

Clustering methods have efficient and fast properties, and one of them is the K-means method, which generates a cluster object based on the attribute to the k partition. This approach works by first determining the cluster that will be produced, then selecting the first member in each cluster to serve as the midpoint (centroid), and then repeating the stages until no object can be moved again [33].

The clustering method which used is K-means. K-means is an algorithm used in the grouping in a manner that separates the data into different groups. This algorithm is able to minimize the distance between the data to the cluster. Basically, the usage of algorithm in the clustering process depends on the data obtained and the conclusion to be achieved at the end of the process [34].

K-means is a widely utilised unsupervised learning technique that addresses the clustering problem. Let X = {x1,...,xn} represent a dataset in a d-dimensional Euclidean space Rd. Let A = {a1, ..., ac} represent the c cluster centres. Define z = [zik]n×c, where zik is a value in binary form (i.e., zik ∈ {0,1}) that signifies whether the information at the point xi is part of the k-th cluster, with k = 1,··· ,c [29].

The k-means objective function is

$$J(z,A) = \sum_{i=1}^{n} \sum_{k=1}^{c} zik \, \|xi - ak\|^2 \quad (3)$$

The k-means algorithm is iteratively applied to satisfy the requisite conditions for minimizing the k-means objective function J(z,A), by changing the formulas for cluster centers and memberships [29], respectively, as follows:

$$ak = \frac{\sum_{i=1}^{n} Zik\ Xij}{\sum_{i=1}^{n} Zik} \quad (4)$$

$$Zik = \begin{cases} 1 \text{ if } \|xi - ak\|^2 = \min_{1 \le k \le c} \|xi - ak\|^2 \\ 0, \text{ otherwise} \end{cases}$$

The Euclidean distance between data point xi and cluster center ak is denoted as $\| xi - ak \|$.

Then, in the below is Table 2 which shown the variables and data that used in this research:

Table 2. Data Variabel Effectiveness and Efficiency Port

Port	RSS	NTS	ND	OC	C	PS
Tanjung Priok	1	1	1	0.866667	0.421053	0.945205
Panjang	1	1	0.4	0.866667	0.263158	0.726027
Tanjung Perak	1	0.444444	1	0.577778	0.052632	1
Teluk Bayur	1	1	0	0.688889	0.052632	1
Banjarmasin	1	0	0.4	0.822222	1	0.534247
Tanjung Emas	1	0.444444	1	0.488889	0	0.972603
Palembang	1	0.444444	0.4	0.622222	0.052632	0.780822
Belawan	1	1	0.4	0.377778	0.263158	0.90411
Balikpapan	1	0	0.4	1	0	0.452055
Samarinda	0	0	0.4	0.688889	0.947368	0.479452
Pontianak	1	0	0.4	0.777778	0.052632	0.39726
Pekanbaru	0	0.444444	0.4	0	0.157895	0.30137
Benoa	0.428571	0	0.4	0.377778	0	0.178082
Tanjung Pinang	0.428571	0.444444	0.4	0.511111	0.105263	0
Lhokseumawe	0	0.444444	0.4	0.2	0.473684	0.616438
Tanjung Intan	1	0.444444	1	0.133333	0	0.547945
Banten	0	1	1	0.133333	0	0.767123
Dumai	0.428571	0.444444	0.4	0.777778	0.947368	0.342466

4. Results and Discussion

4.1. Principle Component Analysis (PCA)

Principal Component Analysis (PCA) is a statistic technique for reduce the number of variables until had several new variables [21]. For the new variables which formed, the interaction between the variables has been eliminated. In this research, there were 6 variables are used to determine the new hub port. The variables which used are regional spatial system, national transportation system, commodity, port service, national defense, operational cost.

These variables were collected from [26]. All of this data had gone through the standardization procedure. In the calculating procedure, the standardization approach is utilized to generate a valuation figure that is comparable among the criteria analyzed using integers ranging from 0 to 1.

Data which had been standardized then will be tested using Barlett Test of Sphericity. The purpose this test is to know whether there is a significant correlation between the variables. If the significance value of Barlett Test of Sphericity test results < 0.05, then there is no correlation between the variables [30]. Besides that, there were also Keiser-Meyer-Oklin (KMO) test was used to measure sample of adequacy (*sampling adequacy*) [31]. The requirement of sample adequacy is the KMO value was obtained greater than or equal to 0.05.

The result of Barlett Test of Sphericity and Keiser-Meyer-Oklin test can be seen in Figure 2. In Barlett Test of Sphericity H_0 which owned is not have a difference. In the Figure 2 of this research can be known that the that the value of Barlett Test of Sphericity is 27.081 with the significant value 0.028. It means this research has a very significant correlation between variables. At the same time, there are also Keiser-Meyer-Oklin test calculation equal to 0.445 or 0.5 which indicates the adequacy that used in this research had been qualified based on the analysis using Principal Component Analysis (PCA).

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.445
Bartlett's Test of Sphericity	Approx. Chi Square df	27.081
		15
	Sig.	.028

Figure 2. Barlett Test of Sphericity and Keiser-Meyer-Oklin Test

After fulfilling both of tests, then Principal Component Analysis can be done. Principle Component Analysis can be done by looking at the eigenvalues in each variable. The new variable (component analysis) which formed based on the eigenvalue should be more than one. The result of eigenvalue and the variety of values calculation can be seen in Table 3.

Table 3. Eigenvalues

Component	Total	Initial Eigenvalues % of Variance	Cumulative %
1	2.117	35.280	35.280
2	1.652	27.536	62.816
3	.958	15.968	78.784
4	.791	13.189	91.973
5	.288	4.802	96.775
6	.194	3.225	100.000

Based on the Table 3 can be known that, there are two new variables which had eigenvalue more than one. In the first principal component had eigenvalue equal to 2.117 with the variance value of 35.28% and the second principal component has an eigenvalue of 1.652 with the variance value of 27.536%. Both of these variables can explain that diversity of the data is 62.816. The factor rotation was performed to determine what type of variables which affect the effectiveness and efficiency of the existing port.

Table 4. Principal Component Matrix

	Component	
	1	2
Port_Services	.859	.150
National_Transportation_System	.700	-.198
National_Defense	.547	-.369
Commodity	-.494	.361
Operational_Cost	.002	.943
Regional_Spatial_System	.589	.660

Extraction Method: Principal Component Analysis. a. 2 components extracted.

In the Table 4 can be seen the correlation between the original variable with the new variable (principal component). The new variables are the result of PCA which then called as loading value. The loading value which used had the value more than 0.5. The following are the variables affecting the port effectiveness and efficiency:

Table 5. Principle Component Result

Principal Component (PC)	Variable	Loading Factor	Variable Explained
PC 1	Port Service National Transportation System National Defense Regional Spatial System	0.859 0.7 0.547 0.589	35.280 %
PC 2	Operational Cost Regional Spatial System	0.943 0.66	27.536

In Table 5 can be known that the first principal component was formed by port service, national transportation system, national defense, and regional spatial system. The first principal component had the variance percentage value of 35.280%. The second principal component has a variance of percentage value of 27.536% which is formed by the operational cost and regional spatial system

4.2. Clustering

This research was grouping into 18 ports based on the result of Principle Component Analysis (PCA) which have been gained 5 variable such as regional spatial system, national transportation system, national defense, operational cost, port services. The result of the cluster data processing can be seen in the Table 6.

Based on the result of clustering K-means calculation, it was obtained the member of Cluster 1 are 7 ports which are Banjamasin, Balikpapan, Samarinda, Pontianak, Benoa, Tanjung Pinang, and Dumai. Then, in the Cluster 2 consist of 7 port which are Tanjung Priok, Panjang, Tanjung Perak, Teluk Bayur, Tanjung Emas, Palembang, and Belawan. Meanwhile, in the Cluster 3 consist of 4 ports which are Pekanbaru, Lhokseumawe, Tanjung Intan. Banten.

Table 6. Result of Clustering Process

Number	Ports	Cluster	Number	Ports	Cluster
1	Banjarmasin	1	10	Tanjung Perak	2
2	Balikpapan	1	11	Teluk Bayur	2
3	Samarinda	1	12	Tanjung Emas	2
4	Pontianak	1	13	Palembang	2
5	Benoa	1	14	Belawan	2
6	Tanjung Pinang	1	15	Pekanbaru	3
7	Dumai	1	16	Lhokseumawe	3
8	Tanjung Priok	2	17	Tanjung Intan	3
9	Panjang	2	18	Banten	3

The respectively of characteristics in the clustering can be seen in the Table 7.

Table 7. Final Cluster Centres

	Cluster		
Regional_Spatial_System	,09	,14	,04
National_Transportation System	,01	,07	,05
National_Defense	,02	,03	,04
Operational_Cost	,33	,30	,06
Port_Services	,29	,70	,45

The characteristics for each cluster can be known in the above Table 7 which can be explained in the below:

1. Cluster 1 had the regional spatial system value in the middle value compared with the other cluster, have the highest value of operational cost compared with the other cluster, and have the lowest value of national transportation system, national defence, also post services compared with the other cluster.

2. Cluster 2 had the highest value of regional spatial system, national transportation system, and port services yang compared with another cluster. In the cluster 2 also had the middle value for national defence and operational cost compared with another cluster.

3. Cluster 3 had the lowest value for regional spatial system and operating cost compared with other cluster, national transportation system and port services was located in the middle value compared with other cluster, the national defence has the highest value compared with another cluster.

Based on the characteristics for each cluster, it can be known if there are 5 variables which used. In the Cluster 2 had an advantage of value to be used as the main port or hub port. In Indonesian ports which included in the Cluster 2 members are Tanjung Priok, Panjang, Tanjung Perak, Teluk Bayur, Tanjung Emas, Palembang, and Belawan

Figure 3. The result of the best main hub port connectivity

The seven ports marked with red stars in Figure 3, consisting of Tanjung Priok, Panjang, Tanjung Perak, Teluk Bayur, Tanjung Emas, Palembang, and Belawan, are members of Cluster 2, which are identified as having potential for ship traffic. Among these, Tanjung Priok and Panjang ports have the potential to serve as central hubs for sea traffic between Java Island and Sumatera Island, while Tanjung Emas and Tanjung Perak ports are positioned to become central hubs for sea traffic between the Northern Coast of Java and South Kalimantan.

In the Sumatera Island, there are several ports such as Palembang, Belawan and Bayur bay that can be used as a potential port for traffic on the Sumatera Island. Port of Palembang has a potential as the centre port for the area in Southern Sumatera and connected among the Sumatera and Western Kalimantan Island. Teluk Bayur port has potential as a centre traffic in the Western part of Sumatera. Belawan port has a potential as the centre port or Northern Sumatera. Meanwhile, by the existence of the centre port in each of the Western region in Indonesia, can increase the effectiveness and efficiency by the existence of the time problem and the sea transportation cost problem.

4.3. Discussion

The study's results demonstrate that the Principal Component Analysis method effectively identifies significant variables and minimises the variables required for Hub Port Connectivity. This study identifies the constituent variables utilised in the clustering process and assesses Hub Port Connectivity, comprising five variables: regional spatial system, national transportation system, national defenses, operational cost, and port services. Divergent from prior research that employed PCA in port service instances. Abbes [24] employed the PCA approach to ascertain the principal determinants of competitiveness within African seaport systems, revealing that handling costs, connectivity, capacity, and the quality of port infrastructure are significant factors in elucidating port competitiveness. Research comparing and measuring the efficiency and competitiveness of the 30 largest container ports in Europe was undertaken by Kammoun and Abdennadher [35].

This study analysed the characteristics derived from the PCA approach, which were subsequently utilised for clustering ports in Indonesia to identify Hub Port connectivity. Bai et al. [14] utilised PCA and compared it with the Liner Shipping connection Index (LSCI), concluding that connection measurement can be enhanced beyond indices that predominantly rely on traffic volume data. This aligns with the study's conclusions about the factors of regional spatial systems, national transportation systems, and port services. This study uses the results of the variables derived from the PCA method as inputs for the clustering technique, aimed at categorising probable ports in Indonesia for Hub Port Connectivity. The study's results indicate that cluster 1 encompasses the ports of Banjarmasin, Balikpapan, Samarinda, Pontianak, Benoa, Tanjung Pinang, and Dumai, which exhibit analogous characteristics. Subsequently, Cluster 2 identifies the correlation among Port Connectivity at Tanjung Priok, Panjang, Tanjung Perak, Teluk Bayur, Tanjung Emas, Palembang, and Belawan. Cluster 3 encompasses Pekanbaru, Lhokseumawe, Tanjung Intan, and Banten. This case study research examines Hub Port Connectivity in Indonesia, distinguishing it from other studies that analyse port connectivity in 29 European ports and apply findings to China and intra-European connection networks [36]. Additionally, the connectivity of the primary Canarian ports is evaluated using graph theory [37], while the characteristics of the maritime shipping network in Northeast Asia are identified [38]. Furthermore, the connectivity of the Asian-Australian cruise shipping network is investigated using the exponential random graph model (ERGM) [25]. This study yielded novel insights and findings pertaining to Hub Port Connectivity in Indonesia.

Moreover, the analysis of cluster 2 member data indicates that Tanjung Priok is suitable as the primary port for Hub Port Connectivity. This conclusion is drawn from the evaluated variables, including regional spatial system, national transportation system, national defenses, operational cost, and port services, with Tanjung Priok achieving the highest score of 4.36. For cluster 1, Banjarmasin was identified as the primary Hub Port Connectivity, achieving a total score of 2.93, while cluster 3 designated Tanjung Intan with a score of 2.99. The score is derived from the aggregation of the total values of each analyzed variable, specifically the regional spatial system, national transportation system, national defence, operational costs, and port services.

We propose policy implications derived from the study on Hub Port Connectivity in Indonesia. The First, prioritize the enhancement of Hub Ports within cluster 2 (Tanjung Priok, Panjang, Tanjung Perak, Teluk Bayur, Tanjung Emas, Palembang, and Belawan). Additionally, given Tanjung Priok's excellent rating and recognized appropriateness as the principal Hub Port for Cluster 2, substantial investment should be allocated to improving its infrastructure, technology, and operating efficiency. For instance, augmenting infrastructure (deepening harbors, expanding terminals, enhancing cargo handling facilities, and modernizing equipment), technological innovations through the adoption of advanced technologies like automation and digitalization [39], [40], [41], and connectivity improvements by enhancing links to inland transportation networks via superior road, rail, and inland waterway connections [42], [43]. Secondly, establish Secondary Hub Ports in Banjarmasin (Cluster 1) and Tanjung Intan (Cluster 3) through investments aimed at enhancing these ports as secondary hubs to bolster regional connectivity and economic development in their respective locales. Third, improve connectivity inside and among clusters by reinforcing transportation links between ports in each cluster to enable efficient cargo movement and minimize logistical expenses. Fourth, guarantee efficient coordination and collaboration among pertinent government agencies, including the Ministry of Transportation, the Ministry of Trade, and the Ministry of Maritime Affairs and Fisheries [44], [45]. Subsequently, consistently assess and analyses the implementation progress and make requisite modifications to guarantee the efficacy of policies. Fifth, promote public-private partnerships to utilize private sector investment and skills in the development and operation of port infrastructure and services. By executing these policy proposals, the Indonesian government may strategically utilize its geographic location to establish a world-class port infrastructure that fosters economic growth, improves regional connections, and bolsters the nation's standing in global trade.

5. Conclusion

The result of principle component analysis can be seen that the first principal component is formed by the service port, national transportation system, national defense, and regional spatial system. The first principal component has a variance percentage value of 35.280%. The second principal component is formed by the operational cost and regional spatial system. The second principal component has a variance percentage value of 27.536%. The result of Principle Component Analysis (PCA) which have been gained 5 variables are regional spatial system, national transportation system, national defense, operational cost, port services. Therefore, it can be formed clustering using K-means which obtained the results of 3 cluster port with each characteristic. There are Cluster 2 which potentially to serve as the main ports in port transportation in Western Indonesia such as Tanjung Priok, Panjang, Tanjung Perak, Teluk bayur, Tanjung Emas, Palembang, and Belawan. The Port of Tanjung Priok and Panjang has the potential to serve as the main port for the marine traffic which occurs between Java Island and Sumatera Island. Port of Tanjung Emas and Tanjung Perak can serve as the port for marine traffic that occurs between North coast of Java and South Kalimantan. Port of Palembang has potential as a centre port for the area in Southern Sumatera and connected among the Sumatera Island and West Kalimantan. Teluk Bayur port has potential as centre traffic in Western Sumatera region. Belawan port also has a potential as the centre port for Northern Sumatera region.

References

[1] M. B. Zaman, I. Vanany, and K. D. Awaluddin, “Connectivity analysis of port in Eastern Indonesia,” *Procedia Earth Planet. Sci.*, vol. 14, pp. 118–127, 2015.
[2] “Home | Logistics Performance Index (LPI).” Accessed: Dec. 19, 2024. [Online]. Available: https://lpi.worldbank.org/
[3] S. Balakrishnan, T. Lim, and Z. Zhang, “A methodology for evaluating the economic risks of hurricane-related disruptions to port operations,” *Transp. Res. Part Policy Pract.*, vol. 162, pp. 58–79, 2022.
[4] Z. Fang, H. Yu, F. Lu, M. Feng, and M. Huang, “Maritime network dynamics before and after international events,” *J. Geogr. Sci.*, vol. 28, no. 7, pp. 937–956, Jul. 2018, doi: 10.1007/s11442-018-1514-9.
[5] W. Li, A. Asadabadi, and E. Miller-Hooks, “Enhancing resilience through port coalitions in maritime freight networks,” *Transp. Res. Part Policy Pract.*, vol. 157, pp. 1–23, 2022.
[6] C. Ferrari, L. Persico, and A. Tei, “Covid-19 and seaborne trade: The Italian perspective,” *Res. Transp. Econ.*, vol. 93, p. 101162, 2022.
[7] L. Rousset and C. Ducruet, “Disruptions in Spatial Networks: a Comparative Study of Major Shocks Affecting Ports and Shipping Patterns,” *Netw. Spat. Econ.*, vol. 20, no. 2, pp. 423–447, Jun. 2020, doi: 10.1007/s11067-019-09482-5.
[8] T. Saito, R. Shibasaki, S. Murakami, K. Tsubota, and T. Matsuda, “Global maritime container shipping networks 1969–1981: Emergence of container shipping and reopening of the Suez Canal,” *J. Mar. Sci. Eng.*, vol. 10, no. 5, p. 602, 2022.
[9] J. Verschuur, E. E. Koks, and J. W. Hall, “Port disruptions due to natural disasters: Insights into port and logistics resilience,” *Transp. Res. Part Transp. Environ.*, vol. 85, p. 102393, 2020.
[10] B. P. S. Indonesia, “Statistik Transportasi Laut 2023.” Accessed: Dec. 19, 2024. [Online]. Available: https://www.bps.go.id/id/publication/2024/11/25/3fcd0493a7b0a2de5a18004b/sea-transportation-statistics-2023.html
[11] J. M. Low, S. W. Lam, and L. C. Tang, “Assessment of hub status among Asian ports from a network perspective,” *Transp. Res. Part Policy Pract.*, vol. 43, no. 6, pp. 593–606, 2009.
[12] J. Y. Kim and Y. Park, “Connectivity analysis of transshipments at a cargo hub airport,” *J. Air Transp. Manag.*, vol. 18, no. 1, pp. 12–15, 2012.
[13] J. I. Castillo-Manzano, L. López-Valpuesta, and D. J. Pedregal, “What role will hubs play in the LCC point-to-point connections era? The Spanish experience,” *J. Transp. Geogr.*, vol. 24, pp. 262–270, 2012.
[14] X. Bai, L. Cheng, D. Yang, and O. Cai, “Does the traffic volume of a port determine connectivity? Revisiting port connectivity measures with high-frequency satellite data,” *J. Transp. Geogr.*, vol. 102, p. 103385, 2022.
[15] A. Asadabadi and E. Miller-Hooks, “Maritime port network resiliency and reliability through co-opetition,” *Transp. Res. Part E Logist. Transp. Rev.*, vol. 137, p. 101916, 2020.
[16] J. Guo, S. Guo, and J. Lv, “Potential spatial effects of opening Arctic shipping routes on the shipping network of ports between China and Europe,” *Mar. Policy*, vol. 136, p. 104885, 2022.
[17] Y. He, Y. Yang, M. Wang, and X. Zhang, “Resilience analysis of container port shipping network structure: The case of China,” *Sustainability*, vol. 14, no. 15, p. 9489, 2022.
[18] L. Kang, W. Wu, H. Yu, and F. Su, “Global container port network linkages and topology in 2021,” *Sensors*, vol. 22, no. 15, p. 5889, 2022.
[19] J. Corey, Q. Wang, J. Zheng, Y. Sun, H. Du, and Z. Zhu, “Container transshipment via a regional hub port: A case of the Caribbean Sea region,” *Ocean Coast. Manag.*, vol. 217, p. 105999, 2022.
[20] D. Tsiotas and C. Ducruet, “Measuring the effect of distance on the network topology of the Global Container Shipping Network,” *Sci. Rep.*, vol. 11, no. 1, p. 21250, 2021.
[21] D. Tocchi, C. Sys, A. Papola, F. Tinessa, F. Simonelli, and V. Marzano, “Hypergraph-based centrality metrics for maritime container service networks: A worldwide application,” *J. Transp. Geogr.*, vol. 98, p. 103225, 2022.
[22] J. Wu, J. Lu, L. Zhang, and H. Fan, “Spatial heterogeneity among different-sized port communities in directed-weighted global liner shipping network,” *J. Transp. Geogr.*, vol. 114, p. 103781, 2024.
[23] Z. Li, H. Li, Q. Zhang, and X. Qi, “Data-driven research on the impact of COVID-19 on the global container shipping network,” *Ocean Coast. Manag.*, vol. 248, p. 106969, 2024.
[24] S. Abbes, “Seaport competitiveness: a comparative empirical analysis between North and West African countries using principal component analysis,” *Int. J. Transp. Econ. Riv. Internazionale Econ. Dei Trasp. XLII 3 2015*, pp. 289–314, 2015.

[25] M. Kanrak and H.-O. Nguyen, “An analysis of connectivity, assortativity and cluster structure of the Asian-Australasian cruise shipping network,” *Marit. Transp. Res.*, vol. 3, p. 100048, 2022.
[26] H. Sutomo and J. Soemardjito, “Assessment Model of the Port Effectiveness and Efficiency (Case Study: Western Indonesia Region),” *Procedia-Soc. Behav. Sci.*, vol. 43, pp. 24–32, 2012.
[27] D. T. Pham and A. A. Afify, “Clustering techniques and their applications in engineering,” *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, vol. 221, no. 11, pp. 1445–1459, Nov. 2007, doi: 10.1243/09544062JMES508.
[28] C. Ding and X. He, “*K* -means clustering via principal component analysis,” in *Twenty-first international conference on Machine learning - ICML ’04*, Banff, Alberta, Canada: ACM Press, 2004, p. 29. doi: 10.1145/1015330.1015408.
[29] K. P. Sinaga and M.-S. Yang, “Unsupervised K-means clustering algorithm,” *IEEE Access*, vol. 8, pp. 80716–80727, 2020.
[30] H. B. Umar, “Principal Component Analysis (PCA) dan aplikasinya dengan SPSS,” *J. Kesehat. Masy. Andalas*, vol. 3, no. 2, pp. 97–101, 2009.
[31] H. W. Oetomo, “Analisis Faktor Ruangan yang Berpengaruh Terhadap Nilai Tanah Perkotaan,” *Econ. J. Emerg. Mark.*, 2006, Accessed: Dec. 19, 2024. [Online]. Available: https://journal.uii.ac.id/JEP/article/view/524
[32] J. O. Ong, “Implementasi algoritma k-means clustering untuk menentukan strategi marketing president university,” 2013.
[33] Y. Ardilla, H. Tjandrasa, and I. Arieshanti, “Deteksi Penyakit Epilepsi dengan Menggunakan Entropi Permutasi, K-means Clustering, dan Multilayer Perceptron,” *J. Tek. ITS*, vol. 3, no. 1, pp. A70–A74, 2014.
[34] S. Agustina, D. Yhudo, H. Santoso, N. Marnasusanto, A. Tirtana, and F. Khusnu, “Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means,” *Univ. Brawijaya Malang Malang*, 2012, Accessed: Dec. 19, 2024. [Online]. Available: https://www.academia.edu/download/46692771/clustering-kualitas-beras-dengan-k-means.pdf
[35] R. Kammoun and C. Abdennadher, “Seaport efficiency and competitiveness in European seaports,” *Transp. Policy*, vol. 121, pp. 113–124, 2022.
[36] Q. Liu, Y. Yang, L. Ke, and A. K. Ng, “Structures of port connectivity, competition, and shipping networks in Europe,” *J. Transp. Geogr.*, vol. 102, p. 103360, 2022.
[37] B. Tovar, R. Hernández, and H. Rodríguez-Déniz, “Container port competitiveness and connectivity: The Canary Islands main ports case,” *Transp. Policy*, vol. 38, pp. 40–51, 2015.
[38] P. N. Nguyen and H. Kim, “Analyzing the international connectivity of the major container ports in Northeast Asia,” *Marit. Bus. Rev.*, vol. 7, no. 4, pp. 332–350, 2022.
[39] D. Sjarifudin and H. Kurnia, “The PDCA approach with seven quality tools for quality improvement men’s formal jackets in Indonesia garment industry,” *J. Sist. Tek. Ind.*, vol. 24, no. 2, pp. 159–176, 2022.
[40] N. Matondang and A. Ishak, “A Literature Review of Distribution Solving Problems by Various Transportation Methods,” *J. Sist. Tek. Ind.*, vol. 25, no. 2, pp. 178–187, 2023.
[41] Y. Prawira, A. Ishak, and A. Anizar, “A Review of Literature on Lean Manufacturing Tools and Implementation Based on Case Studies,” *J. Sist. Tek. Ind.*, vol. 26, no. 1, pp. 11–21, 2024.
[42] X. Zhu, *Smart Road Infrastructure: Ideas, Innovations and Emerging Technologies*. in Lecture Notes in Intelligent Transportation and Infrastructure. Singapore: Springer Nature Singapore, 2024. doi: 10.1007/978-981-97-3831-1.
[43] R. Krishnan and Y. Shukla, “Use of geospatial technologies in transport infrastructure”, Accessed: Dec. 24, 2024. [Online]. Available: https://www.researchgate.net/profile/Yogita-Shukla/publication/292971616_Innovative_Geospatial_Solutions_-_Key_to_India%27s_Transport_Infrastrcuture/links/56b2f25608aed7ba3fedf0fd/Innovative-Geospatial-Solutions-Key-to-Indias-Transport-Infrastrcuture.pdf
[44] M. Gómez-Ballesteros *et al.*, “Transboundary cooperation and mechanisms for Maritime Spatial Planning implementation. SIMNORAT project,” *Mar. Policy*, vol. 127, p. 104434, 2021.
[45] H. Sandee, “Improving Connectivity in Indonesia: The Challenges of Better Infrastructure, Better Regulations, and Better Coordination,” *Asian Econ. Policy Rev.*, vol. 11, no. 2, pp. 222–238, Jul. 2016, doi: 10.1111/aepr.12138.