The current issue and full text archive of this journal is available on International Journal of Economics and management review (IJEMR) Insight at: https://jurnal.smartindo.org/index.php/ijemr

Unraveling The Supply-Side Factors Shaping East Java's Economy: Insights From PCA and Machine Learning

Muhammad Jamie Rofie Quality
Department of Economics, University of Indonesia
Muhammad Firdaus Al Farohi
Department of Accounting, University of Indonesia
As'ary Ricklas Hidayat
Department of Economics, University of Indonesia

Abstract

Purpose - Examining the determinant factors from the production or supply side that affect the economic performance of East Java.

Design/methodology/approach – By using data from the Central Statistics Agency (BPS) regarding the factors supporting economic growth from the production side, this research aims to examine the determinant factors that affect the economic performance of East Java. Through Machine Learning analysis using principal component analysis and clustering analysis, certain characteristics were found among districts and cities in East Java.

Originality - Analyzing the economic structure of East Java Province specifically from the demand side

Findings and Discussion – PCA was used to reduce the number of variables and resulted in several components that are consistent with general categorization. Urban areas consistently exhibit high human resource components, while another cluster shows high dependence on natural resources.

Conclusion – The intense competition in the market, aftershocks of the pandemic, extreme weather conditions, and rapid social, economic, and technological changes have made the global economic situation much more unstable. This has resulted in economic downturns in various countries. Nevertheless, the Indonesian economy has shown strong growth.

Keywords - Principal Component Analysis, Clustering, Regional Economics, Structural Economics

Introduction

The year 2023 is proving to be a crucial moment for the economies of countries around the world, Indonesia included. The Economist's report, 'The World Ahead 2023,' provides an overview of a world that has become far more unstable due to intense competition among major powers in the market, aftershocks of the pandemic, extreme weather conditions, and rapid social, economic, and technological changes (Wahid, 2022).

This has been a contributing factor to the decline in economic performance in some countries. However, Indonesia stands in contrast to the economic conditions in some countries, as its economic growth, based on Gross Domestic Product (GDP), reached 5.03% in the first quarter (January-March) of 2023, marking an increase of 0.39% compared to the 5.01% in the fourth quarter of 2022 (BPS, 2023). This increase can be attributed to the contributions and economic growth in various regions of Indonesia, with the largest contribution coming from the Regional Gross Domestic Product (RGDP) of Java, which accounted for 57.17% in Q1 2023, followed by Sumatra (21.82%), Kalimantan (9%), Sulawesi (6.87%), Bali and Nusa Tenggara (2.68%), and Maluku and Papua (2.46%) (BPS, 2023). Among these eight islands, several provinces have made significant contributions to Indonesia's GDP. East Java, as one of the provinces in Indonesia, contributed 14.26% to Indonesia's GDP. East Java's RGDP contribution to Indonesia's GDP is the second-largest after DKI Jakarta, which stands at 16.64% (BPS, 2023).

The highly strategic geographical position makes East Java one of the vital hubs for economic growth in Indonesia, particularly for the eastern part of the country (Rasyid, 2016). Economic growth in East Java in the fourth quarter of 2022 increased compared to the previous year's fourth quarter. In the quarterly report, East Java's economic performance grew by 5.34% (YoY), compared to 3.56% (YoY) in 2021. This improvement can be attributed to various factors, such as increased investments that have led to the creation of new capital, thereby absorbing new production factors, like job creation and reduced unemployment rates (Rochaida, 2022). Furthermore, this economic improvement aligns with national and Java regional economic trends, viewed from both the supply and demand sides. The improvement in economic performance from the demand side is driven by higher growth in Household Consumption, Government Consumption, and Gross Fixed Capital Formation. On the supply side, the improvement is driven by higher growth in Manufacturing, Transportation and Warehousing, Accommodation and Food Services, Financial and Insurance Services, Real Estate, Business Services, Government Administration, Agriculture, and Mandatory Social Security and Other Services (BI, 2023).

Given these observations, this research will attempt to examine the determinants from the production or supply side that influence the economic performance of East Java. In conducting this research, the researcher will use data that describes the aspects supporting economic growth in East Java.

Literature Review

Economic growth is the primary goal desired by every government to ensure the improvement of the welfare of its citizens. This is one of the indicators used to determine the success of a government. There are various theories that can serve as the foundation for examining the level of economic growth. One well-known economic growth theory is the Keynesian theory, stemming from Keynes' General Theory (1937), which explains that economic output is influenced by consumption, government spending, and investment. Another commonly used theory is the Endogenous Growth Model proposed by Romer (1989), which states that output is influenced by capital, labor, human capital, and the rate of technological growth. If we refer to the Solow Growth Model developed by Solow (1956), it describes that economic growth is significantly affected by the level of technology, capital accumulation, and labor. These three aspects interact with each other and ultimately affect a country's economic output. Overall, all growth theories emphasize that factors such as capital, technology, labor, and human capital are crucial aspects of economic growth. These four factors differentiate the economic growth of regions, both at the national and regional levels.

A. Economic Structure

Research conducted by Aginta and Someya (2022) explains that the dynamics of economic growth are significantly influenced by economic structure, which can be measured by the proportion of specific industrial sectors, human capital, infrastructure development, and other indicators. The dominance of specific economic sectors can ultimately affect a region's economic resilience in response to monetary policy transmission. In another study, the development of industrial structure has proven to drive massive regional economic development. Chen et al. (2021) show that regional industrial transitions and adjustments can be a solution for regional economic growth. Besides driving economic growth, the transformation and transition of economic structures can lead to regional convergence in an area influenced by similar aspects such as geographical, social, and industrial conditions (Abdulla, 2021). Analyzing economic structure can reveal the economic identity of a region, which can then be used to understand its economic potential. This, in turn, encourages regional convergence based on similarities in conditions, making it easier to determine the characteristics of relationships and needs between regions. Although doubts have arisen about the impact of strengthening economic structures through industrial structure reinforcement on regional resilience in facing economic shocks, such as those caused by the COVID-19 pandemic, Kim, Lim, and Colletta (2022) state that the industrial structure in a region, whether essential or non-essential, high-interaction or low-interaction, does not determine the economic stability of the region when facing economic shocks. Instead, government policies and the quality of human resources, as reflected in compliance levels, have a more significant impact.

B. Regional Development

Regional development is a priority for local governments to improve the welfare of their citizens. An important aspect of regional development is regional education, including improving the quality of education and the skills of the regional population, particularly entrepreneurs (Gennaioli et al., 2013). Rodríguez-Pose (2013) explains that another influential factor in regional development is the institutions in the region, which have an impact on determining access to improving the quality of life, such as education and the protection of rights. Meanwhile, the proportion of human capital, the number of creative workers, and technology will drive progress in regional development (Florida, Mellander, and Stolarick, 2008).

Through existing literature, we can understand that economic development must consider regional economic structure. Unfortunately, there is still no research that attempts to analyze the economic structure of a specific region in Indonesia and focuses too much on the analysis of economic structure from the demand side. Furthermore, some existing literature uses regression methods for analysis. Therefore, there is a need for an understanding of regional economic structure from the supply side with different analytical methods to obtain a more holistic understanding of regional economic growth.

Methods, Data, and Analysis

The data used in this research is sourced from a dataset covering 38 regencies/cities in East Jawa issued by the Central Statistics Agency (Badan Pusat Statistik). All the data used are quantitative in nature. Data collection was carried out by combining various types of available data, resulting in a total of 20 variables. The collected data was then interpreted as a variable related to Human Capital, Natural Resources, Capital, Infrastructure, and Output. The data used in this research are secondary data selected based on the Solow Growth Model. In total, there are 760 sample data points from the 38 regencies/cities in East Jawa. The gathered data was processed using Machine Learning with the Principal Component Analysis (PCA) method and clustering using the STATA application. The research process was carried out in the following steps: data mining, data cleaning & standardization, KMO Statistics Analysis, PCA, Clustering, K-Means & Hierarchical Analysis, and Interpretation.

A. Kaiser-Meyer-Olkin (KMO) Statistics

Before conducting data analysis using the PCA and Clustering methods, it is essential to test the adequacy of the sample that has been taken. This test is performed by measuring the correlation between variables to determine the necessity of conducting factor analysis. If the KMO value is less than 0.50 or 0.60, it signifies significance and rejects the hypothesis that factor analysis should be performed using the related data.

B. Principal Component Analysis (PCA)

PCA is used to reduce dimensionality without reducing the variability of the data obtained from the 38 regencies/cities in East Java. By combining several variables linearly, new components are created that will replace the functions of variables in the original data.

C. Clustering

Next, after PCA, clustering is performed on the obtained data to assess the similarity of conditions in each regency/city based on the predetermined variables. Clustering will result in certain subsets referring to the similarity of some variables used in the data. This analysis method is conducted using machine learning to generate specific groups or clusters.

D. K-Means

In the clustering process, there are several methods that can be used, one of which is the K-Means method. This method is classified as partitional clustering, where centroids are assigned, and clusters are created iteratively until convergence is achieved.

E. Hierarchical Clustering

Apart from K-Means, there is a hierarchical clustering method aimed at grouping data progressively based on the distance between data points using linkage methods.

F. Complete Linkage

The formula used in this type of data linkage determines the maximum difference within each cluster. The notation for Complete

Linkage is:

max {d(x,y):x ϵ A, y ϵ B}

G. Ward's Method

In addition to Complete Linkage, grouping can also be determined by referring to the total sum of squares (SSE) within a cluster to determine the resulting groups. The notation for Ward's Method is:

$$SSE = \sum_{x=1}^{K} \sum_{y=1}^{n_i} \left(y_{xy} - \bar{y}_x\right)^2$$

Results

A. Principal Component Analysis (PCA)

Factor analysis is a prerequisite for the sample to be considered adequate for clustering analysis. The results of the analysis using Kaiser-Meyer-Olkin (KMO) are presented in Table 2. An overall KMO value above 0.7 indicates that factor analysis can be performed. Principal Component Analysis (PCA) is conducted to reduce the variables used as determinants into distinct components. Before conducting PCA estimation, the correlation between variables is analyzed to determine if rotation is necessary, as well as the type of rotation to be applied to the sample used. The results of the correlation between variables are shown in Table 3. Based on this table, it can be observed that the majority of variables have low to moderate correlations with each other. Only a few variables, such as avg_school with ipm and forestry with land area, exhibit relatively high correlations, indicating their independence from each other. Therefore, orthogonal rotation using varimax is considered to enhance the interpretability of the model to be used.

In many ways, it is the most important section in an article. Because it is the last thing a reader sees, it can have a major impact on the reader's perceptions of the article and the research conducted. Different authors take different approaches when writing the discussion section, the discussion section should:

- Restate the study's main purpose
- Reaffirm the importance of the study be restating its main contributions
- Summarize the results in relation to each stated research objective or hypothesis without introducing new material

- Relate the findings to the literature and the results reported by other researches
- Provide possible explanations for unexpected or non-significant findings
- Discuss the managerial implications of the study
- Highlight the main limitations of the study that could influence its internal and external validity
- Discuss insightful (i.e., non-obvious) directions or opportunities for future research on the topic

Table 1: The Result of KMO Adequacy Test

Kaiser-Meyer-Olkin measure of sampling adequacy	
Variable	kmo
ah_hidup_std	0.7227
avg_school_std	0.8086
ipg_std	0.8852
ipm_std	0.8056
pov_rate_std	0.8113
gender_ratio_std	0.5490
res_dist_std	0.5904
Gini_std	0.6598
unemp_std	0.7456
harvest_std	0.6886
plantation_std	0.8491
forestry_std	0.5835
fishing_std	0.6137
int_speed_std	0.7048
konstruksi_std	0.2916
luas_wil_std	0.7298
fourg_vil_std	0.8136
prov_road_std	0.7288
dmg_road_std	0.3051
Overall	0.7332

Source: Author's Compilation

Table 2: The Correlations between Variables in the sample

	life_exp	avg_sc	ipg	ipm	pov_ra	gender	res_dis	Gini	unemp	harvest	planta~	forestr	fishing	int_s~	konstr~	luas_w	fourg_~	prov_~	dmg_ro
life_exp	1.00																		
avg_school	0.72	1.00																	
ipg	0.59	0.73	1.00																
ipm	0.74	0.98	0.72	1.00															
pov_rate	-0.58	-0.82	-0.69	-0.81	1.00														
gender_ratio	0.44	0.25	0.30	0.23	-0.34	1.00													
res_dist	0.00	-0.18	-0.29	-0.10	0.04	0.27	1.00												
Gini	0.38	0.44	0.41	0.47	-0.55	0.17	0.07	1.00											
unemp	0.39	0.72	0.49	0.74	-0.56	0.21	0.19	0.31	1.00										
harvest	-0.18	-0.46	-0.34	-0.40	0.36	0.16	0.40	-0.20	-0.29	1.00									
plantation	-0.47	-0.60	-0.72	-0.57	0.36	-0.24	0.38	-0.04	-0.51	0.35	1.00								
forestry	-0.33	-0.36	-0.42	-0.32	-0.03	0.04	0.45	0.18	-0.18	0.33	0.67	1.00							
fishing	-0.14	-0.31	-0.50	-0.27	0.42	-0.25	0.19	-0.35	-0.16	0.35	0.33	0.10	1.00						
int_speed	0.20	0.37	0.12	0.36	-0.34	0.20	0.10	0.14	0.34	-0.12	-0.03	0.15	-0.31	1.00					
konstruksi	0.05	0.07	-0.06	0.16	-0.03	-0.05	0.62	0.02	0.29	-0.08	0.07	0.00	0.07	-0.01	1.00				
luas_wil	-0.31	-0.54	-0.59	-0.49	0.27	0.11	0.56	0.05	-0.35	0.60	0.73	0.84	0.36	-0.03	0.05	1.00			
fourg_vil	0.49	0.77	0.60	0.75	-0.79	0.29	0.00	0.46	0.63	-0.20	-0.31	-0.12	-0.23	0.32	-0.01	-0.29	1.00		
prov_road	-0.29	-0.39	-0.49	-0.39	0.16	0.29	0.48	-0.03	-0.31	0.43	0.51	0.49	0.15	0.03	0.07	0.59	-0.17	1.00	
dmg_road	-0.07	0.00	0.08	-0.04	-0.11	0.17	-0.06	-0.04	-0.09	0.02	-0.01	-0.01	-0.01	-0.16	-0.04	-0.01	0.12	0.09	1.00

Source: Author's Compilation

PCA without rotation and PCA with varimax rotation were performed. The results of PCA without rotation are shown in Table 4. The eigenvalues in PCA are illustrated in Figure 1. Based on theeigenvalues obtained in PCA, the optimal number of components for the model can also be determined: components with eigenvalues above 1. Therefore, in this analysis, six components are utilized. Using these six components, the model can account for 80.76% of the variability in the existing data, as indicated by the cumulative proportion value. The presentation of PCA loadings has excluded components above 6 and left empty loadings below 0.3, which has become a rule of thumb for sufficiently influential loading values on the respective components.

Table 3: The Eigenvalues for Each PCAs and Their Proportion to the Observations

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	7.1305	4.0240	0.3753	0.3753
Comp2	3.1065	1.5036	0.1635	0.5388
Comp3	1.6028	0.2173	0.0844	0.6231
Comp4	1.3856	0.3061	0.0729	0.6961
Comp5	1.0795	0.0396	0.0568	0.7529
Comp6	1.0398	0.1581	0.0547	0.8076
Comp7	0.8817	0.2823	0.0464	0.8540
Comp8	0.5995	0.1030	0.0316	0.8856
Comp9	0.4965	0.0603	0.0261	0.9117
Comp10	0.4362	0.0765	0.0230	0.9347
Comp11	0.3596	0.0850	0.0189	0.9536
Comp12	0.2746	0.0951	0.0145	0.9680
Comp13	0.1795	0.0358	0.0094	0.9775
Comp14	0.1437	0.0408	0.0076	0.9850
Comp15	0.1029	0.0245	0.0054	0.9905
Comp16	0.0784	0.0238	0.0041	0.9946
Comp17	0.0546	0.0177	0.0029	0.9975
Comp18	0.0369	0.0254	0.0019	0.9994
Comp19	0.0114	.	0.0006	1.0000

Source: Author's Compilation

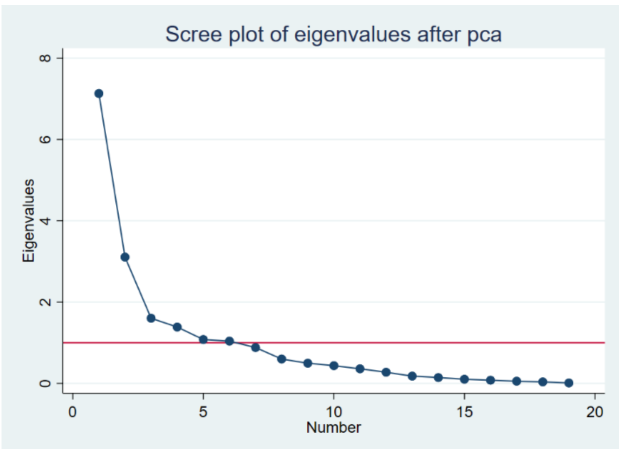

Figure 1: The Eigenvalues (vertical) for Each PCAs Based on the Number of Component(s) (horizontal)

Furthermore, PCA was conducted with orthogonal rotation using the varimax option. This is commonly done for variables assumed to be independent or have relatively low correlations with each other. The results of this rotation are shown in Table 5. There are some changes in the results compared to the previous PCA results. Some variables have changed their component placements due to altered loading values. Since PCA with varimax rotation has only one cross-loading, which is fewer than the regular PCA results, PCA with varimax rotation is used for cluster analysis.

Table 4: The Loadings of PCAs with Varimax Rotation

Variable	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6
life_exp						
avg_school	0.3758					
ipg						
ipm	0.3871					
pov_rate	-0.401					
gender_ratio			0.688			
res_dist				0.5287		
Gini	0.3611	0.3201				
unemp						
harvest			0.4162		0.3558	
plantation		0.4439				
forestry		0.5395				
fishing					0.7451	
int_speed					-0.382	-0.486
konstruksi				0.7628		
luas_wil		0.4298				
fourg_vil	0.4004					
prov_road			0.3308			
dmg_road						0.8064

Source: Author's Compilation

Despite the rotation, there are still issues with the observed data loadings. Three variables exhibit cross-loading, meaning they have loadings above 0.3 on more than one component. In this case, the magnitude of the loading is taken into account to determine the appropriate component for the respective variable. The variable Gini is assigned to Component 1, the variable harvest is assigned to Component 3, and the variable int_speed is assigned to Component 6. Additionally, there are variables that do not have loadings above 0.3 on any of the six components and are therefore not considered in the clustering.

Interpretation of Components:
The six components formed based on the loadings of each variable are as follows:

- Component 1: avg_school, ipm, pov_rate, Gini, fourg_vil

- Component 2: plantation, forestry, luas_wil
- Component 3: gender ratio, harvest, prov_road
- Component 4: res_dist, konstruksi
- Component 5: fishing
- Component 6: int_speed, dmg_road

The grouping of variables based on components reveals that some components have related variables, while others do not, in line with theoretical categorization.

Component 1 portrays the quality of human capital, except for the variable Gini, which is associated with demographic factors, and fourg_vil, which pertains to capital conditions. Component 2 comprises variables related to natural resources. Variables in Component 3, in conjecture, exhibit less consistency, where gender_ratio is related to demographics, harvest is related to natural resources, and prov_road is one of the indicators of infrastructure. Component 4 consists of res_dist, which is a demographic indicator, and konstruksi, which is an indicator of infrastructure progress. Component 5 only contains one variable. Component 6 consists of int_speed, which is an indicator of capital condition, and dmg_road, which is an indicator of infrastructure.

B. Clustering Analysis

1. K-Means Clustering

The first clustering was conducted using the K-Means method, utilizing Euclidean distance as the basis for cluster formation. The results of this method are shown in Table 6. The number of clusters used can be determined using the within-clusters sum of squares (WSS), which can be obtained through the elbow method. By using the eigenvalue graph in Figure 1, it can be determined that the "elbow" point is located at the 3rd cluster. Therefore, 3 clusters are employed for the subsequent analysis. Characteristics using the averages of each component are presented in Table 7. Visualization of the effects of each component is depicted in Figure 2

Table 5: The Result of K-Means Clustering

Kabupaten/Kota	Klaster	Kabupaten/Kota	Klaster
Kab. Pacitan	3	Kab. Magetan	1
Kab. Ponorogo	1	Kab. Ngawi	3
Kab. Trenggalek	1	Kab. Bojonegoro	3
Kab. Tulungagung	1	Kab. Tuban	3
Kab. Blitar	1	Kab. Lamongan	3
Kab. Kediri	1	Kab. Gresik	1
Kab. Malang	3	Kab. Bangkalan	2
Kab. Lumajang	3	Kab. Sampang	2
Kab. Jember	3	Kab. Pamekasan	2
Kab. Banyuwangi	3	Kab. Sumenep	2
Kab. Bondowoso	2	Kota Kediri	1
Kab. Situbondo	2	Kota Blitar	1
Kab. Probolinggo	2	Kota Malang	1
Kab. Pasuruan	3	Kota Probolinggo	1
Kab. Sidoarjo	1	Kota Pasuruan	1
Kab. Mojokerto	1	Kota Mojokerto	1
Kab. Jombang	1	Kota Madiun	1
Kab. Nganjuk	1	Kota Surabaya	1
Kab. Madiun	1	Kota Batu	1

Source: Author's Compilation

Table 6: The Characteristics of Each K-Means Cluster Based on Their Mean Value of Principal Components

	pc1	pc2	pc3	pc4	pc5	pc6
1	1.6234	-1.0057	0.1119	-0.0975	-0.5797	-0.0627
2	-3.3266	0.5154	-1.7865	0.1513	0.6202	-0.1013
3	-1.0806	1.7512	1.0156	0.0989	0.7832	0.2025
Total	3.14E-09	-6.67E-09	9.02E-09	-8.16E-09	2.55E-09	-3.92E-10

Source: Author's Compilation

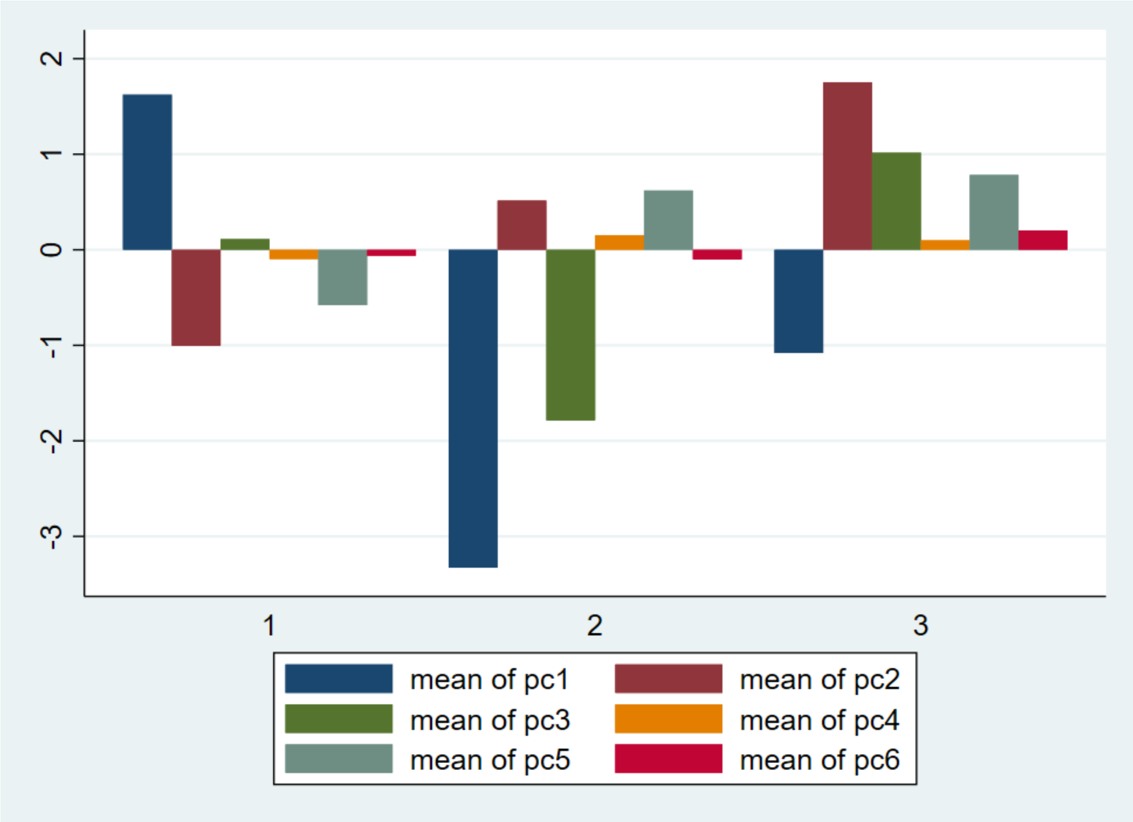

Figure 2: The Characteristics of Each K-Means Cluster Based on Their Mean Value of Principal Components
Source: Author's Compilation

2. Hierarchical Clustering

Clustering analysis using the Hierarchical clustering method offers several linkage options. For practicality, only two of these options are employed: complete linkage and Ward's linkage. The first option is chosen to minimize the effects of outlier data. The second option is selected because it can generate the most homogeneous clusters. Similar to the previous method, we are using 3 clusters.

The results of hierarchical clustering using complete linkage are presented in Figure 3. In the top graph, the clusters are divided until the branches leading to each regency/city are discernible. In the bottom graph, the division into 3 clusters is shown. The majority of regencies/cities belong to Cluster 1 (G1). Cluster 2 (G2) comprises only three regencies: Malang, Banyuwangi, and Jember. All cities and some remaining regencies are grouped in
Cluster 3.

From the table and figures, it can be observed that Cluster 1, which mostly consists of cities, is predominantly influenced by Component 1, as previously mentioned, representing human resource quality. The positive average values indicate that Cluster 1 relatively possesses high-quality human resources. On the other hand, Cluster 2 is dominated negatively by Component 1 and Component 3, indicating a deficiency of variables in these components. Cluster 3 is also dominated by Components 1, 2, and 3, with each having a negative, positive, and positive impact, respectively. Looking at the variables that make up Component 2, it can be concluded that Cluster 3 is relatively more dependent on natural resources compared to the other clusters.

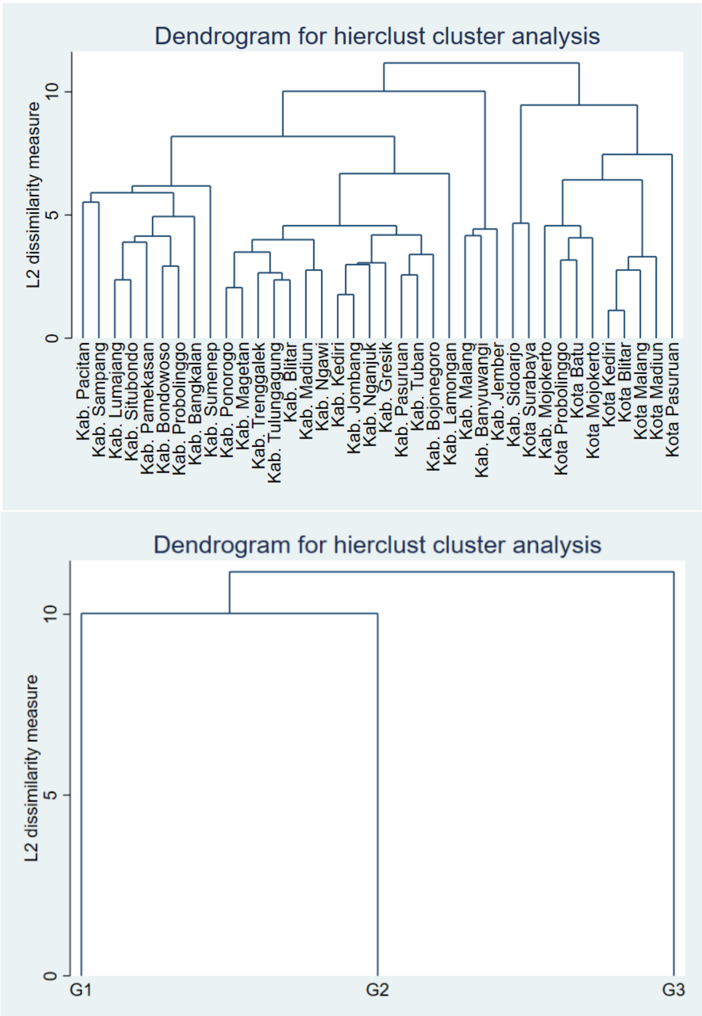

Figure 3: The Dendrogram of Cities and Regencies using Complete Linkage Hierarchical Clustering (top) and The Division into Three Clusters (bottom)
Source: Author's Compilation

Next is the hierarchical clustering analysis using Ward's linkage, as shown in Figure 4. There is a quite significant difference between the results obtained using these two linkage methods. The differences in cluster results are summarized in Table 8, where "hier_com" represents the result of hierarchical clustering with complete linkage, and "hier_ward" represents hierarchical clustering with Ward's linkage. The characteristics of each cluster along with the linkage methods are shown in Table 9 and Figure 5. It's important to note that the cluster numbering generated by the K-Means and hierarchical methods can be quite different due to the arbitrary or random nature of the machine learning process.

Table 7: The Result of Hierarchical Clustering

Kabupaten/Kota	hier_com	hier_ward	Kabupaten/Kota	hier_com	hier_ward
Kab. Pacitan	1	1	Kab. Magetan	1	1
Kab. Ponorogo	1	1	Kab. Ngawi	1	1
Kab. Trenggalek	1	1	Kab. Bojonegoro	1	1
Kab. Tulungagung	1	1	Kab. Tuban	1	1
Kab. Blitar	1	1	Kab. Lamongan	1	1
Kab. Kediri	1	1	Kab. Gresik	1	1
Kab. Malang	2	1	Kab. Bangkalan	1	2
Kab. Lumajang	1	2	Kab. Sampang	1	2
Kab. Jember	2	1	Kab. Pamekasan	1	2
Kab. Banyuwangi	2	1	Kab. Sumenep	1	2
Kab. Bondowoso	1	2	Kota Kediri	3	3
Kab. Situbondo	1	2	Kota Blitar	3	3
Kab. Probolinggo	1	2	Kota Malang	3	3
Kab. Pasuruan	1	1	Kota Probolinggo	3	3
Kab. Sidoarjo	3	3	Kota Pasuruan	3	3
Kab. Mojokerto	3	1	Kota Mojokerto	3	3
Kab. Jombang	1	1	Kota Madiun	3	3
Kab. Nganjuk	1	1	Kota Surabaya	3	3
Kab. Madiun	1	1	Kota Batu	3	3

Source: Author's Compilation

Table 8: The Characteristics of Each Hierarchical Cluster Based on Their Mean Value of Principal Components

Hierarchical Clustering Complete Linkage

	pc1	pc2	pc3	pc4	pc5	pc6
1	-1.2032	0.0454	-0.0389	-0.1978	0.3808	0.1229
2	-0.5209	5.1656	1.4663	1.0639	0.4770	-0.2689
3	2.7673	-1.5079	-0.3151	0.1415	-0.9608	-0.1947
Total	3.14E-09	-6.67E-09	9.02E-09	-8.16E-09	2.55E-09	-3.92E-10

Hierarchical Clustering Ward's Linkage

	pc1	pc2	pc3	pc4	pc5	pc6
1	-0.2534	0.5801	0.9158	-0.1125	0.2364	0.0109
2	-3.1456	0.5603	-1.6082	0.0697	0.5235	0.0685
3	3.0232	-1.6085	-0.5451	0.1693	-0.8917	-0.0765
Total	3.14E-09	-6.67E-09	9.02E-09	-8.16E-09	2.55E-09	-3.92E-10

Source: Author's Compilation

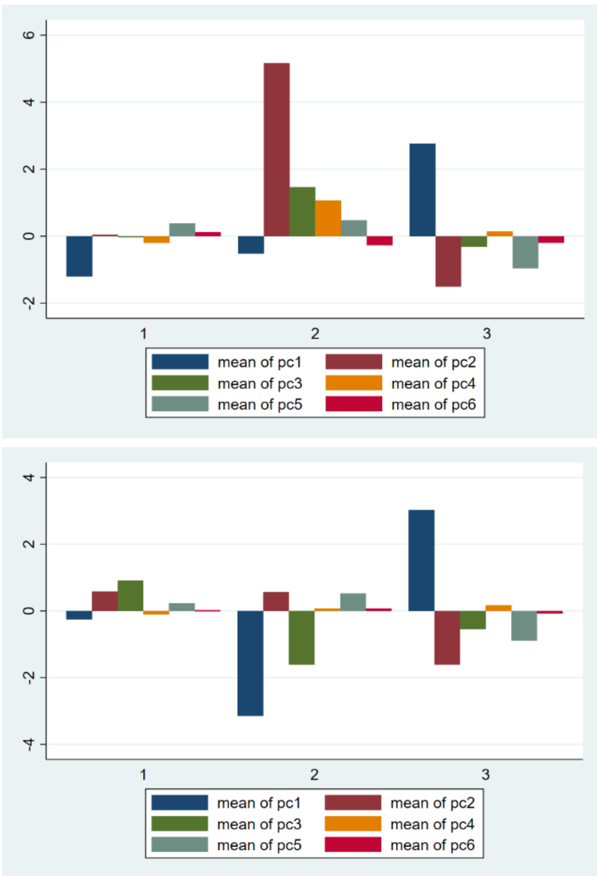

Figure 4: The Characteristics of Complete Linkage (top) and Ward's Linkage Hierarchical Clustering Based on Their Mean Value of Principal Components

Based on the tables and graphs from the hierarchical clustering, several conclusions can be drawn. Both linkage methods yield different cluster members and different characteristics of components. However, in both methods, all cities belong to the same cluster, Cluster 3. Cluster 3 has a dominant and positively valued Component 1 characteristic. In complete linkage, it is evident that Cluster 2 exhibits a relatively high Component 2 characteristic, indicating a dependence on natural resources. A noticeable difference can be observed in the results of Ward's linkage for Cluster 2, where the average value of Component 2 is not as high.

Discussion

The primary objective of this study was to apply clustering analysis, utilizing both K-Means and Hierarchical clustering methods, to categorize cities and regencies in East Java, Indonesia, based on their economic indicators. Additionally, Principal Component Analysis (PCA) was utilized to reduce the dimensionality of the dataset. This research was conducted as part of an economic modelling paper competition, aiming to provide actionable insights for policy-makers and planners in East Java.
The K-Means clustering method provided valuable insights into the grouping of cities and regencies. By employing the elbow method, three clusters were determined as optimal for this analysis. The clusters revealed distinct characteristics:

- Cluster 1 (G1): This cluster predominantly consists of cities and is strongly influenced by Component 1, representing human capital quality. The positive average values signify a relatively high level of human resource quality.
- Cluster 2 (G2): Comprising Malang, Banyuwangi, and Jember, this cluster is characterized by negative influences from both Component 1 and Component 3. This indicates a deficiency in variables related to these components.
- Cluster 3 (G3): This cluster is characterized by a mix of positive and negative influences from Components 1, 2, and 3. Notably, the inclusion of Component 2 variables indicates a relatively higher dependency on natural resources compared to the other clusters.

Two different linkage methods, complete linkage and Ward's linkage, were employed in the hierarchical clustering analysis. The choice of linkage method significantly impacted the resulting clusters. While both methods yielded different cluster members and characteristics, it was noteworthy that all cities consistently fell into Cluster 3 in both cases. This cluster was characterized by a dominant and positively valued Component 1.

In complete linkage, Cluster 2 exhibited a relatively high

Component 2 characteristic, indicating a strong reliance on natural resources. However, a striking difference was observed in the results of Ward's linkage, where the average value of Component 2 in Cluster 2 was not as pronounced. The application of two distinct linkage methods, complete linkage and Ward's linkage, allowed for a nuanced examination of regional patterns. Complete linkage, chosen for its ability to mitigate the influence of outlier data, revealed a clear distinction in Cluster 2, emphasizing the significance of natural resource dependencies. This finding holds implications for regional development strategies, highlighting the need for sustainable resource management. In contrast, Ward's linkage, known for producing the most homogenous clusters, provided a different perspective. The muted influence of Component 2 in Cluster 2 suggests a more balanced economic profile. This insight could be pivotal in guiding policies aimed at diversifying economic activities and enhancing overall stability within these regions. The observed discrepancies between the two linkage methods underscore the importance of considering multiple clustering approaches to gain a comprehensive understanding of regional dynamics.

The comparison between K-Means and Hierarchical clustering methods highlights the importance of considering different clustering techniques. The choice of method can lead to varying interpretations of the data. For instance, the inclusion of Malang, Banyuwangi, and Jember in a distinct cluster (G2) in K-Means analysis was not mirrored in hierarchical clustering, indicating the sensitivity of clustering outcomes to the chosen algorithm.

The findings of this study provide valuable insights for policy-makers and planners in East Java, allowing for targeted interventions based on the characteristics of each cluster. Additionally, the methodological approach employed here can serve as a template for similar studies in other regions.

Conclusion

The analysis of the supply side in the economies of districts and cities in East Java is closely related to the capacity and stock of production factors in each region. This research aims to determine the characteristics of regions with variables that serve as indicators of these production factors using one of the machine learning algorithms, namely clustering analysis and principal component analysis. Based on several clustering methods used, including K-Means, hierarchical complete linkage, and hierarchical Ward's linkage, differences in members were found in each cluster. The main characteristic observed is that urban areas have a unique characteristic of having a high human resource component, while another cluster is highly dependent on natural resources. Based on these findings, there is a need for government policy focus that adjusts to the characteristics of the economic structure among regions by adapting the level of economic dependency, focusing on the development of service industries in regions with a high dependency on human resources and focusing on primary industry development and processing in regions with abundant natural resources.

Limitation

While this study provides valuable insights into the clustering of cities and regencies in East Java, Indonesia, it is essential to acknowledge several inherent limitations. Firstly, the choice of economic indicators significantly influences the clustering outcomes, with alternative indicators potentially leading to different cluster assignments. Additionally, the results are sensitive to the choice of clustering methods (K-Means, Hierarchical), introducing subjectivity into the analysis. The use of Principal Component Analysis (PCA) for dimensionality reduction assumes accurate representation of the underlying economic structure, but variations in component selection may lead to altered cluster assignments. The study predominantly focuses on economic indicators, potentially overlooking spatial patterns that may play a vital role in regional development. Incorporating spatial analysis could offer a more comprehensive understanding of regional dynamics. Furthermore, it is important to note that cluster numbering can vary arbitrarily between different methods (K-Means, Hierarchical), potentially influencing cluster interpretation. The clustering results are contingent on the quality and availability of the data, and any inaccuracies or limitations in the dataset may affect the robustness of findings. Lastly, the insights derived from this study are specific to East Java, Indonesia, and may not be directly applicable to other regions or countries, highlighting the importance of considering regional context in similar analyses.

Reference

Abdulla, K. (2021). Regional convergence and structural transformation in a resource-dependent country. Structural Change and Economic Dynamics, 59, 548-557. https://doi.org/10.1016/j.strueco.2021.10.004

Aginta, H., & Someya, M. (2022). Regional economic structure and heterogeneous effects of monetary policy: Evidence from indonesian provinces. Journal of Economic Structures, 11(1), 1-25. https://doi.org/10.1186/s40008-021-00260-6

Badan Pusat Statistik. (20 C.E., May). Badan Pusat Statistik. Www.bps.go.id. https://www.bps.go.id/pressrelease/2023/05/05/1998/ekonomi-indonesia-triwulan-i-2023-tumbuh-5-03-persen--y-on-y-.html

Badan Pusat Statistik. (2023). Badan Pusat Statistik. Www.bps.go.id. https://www.bps.go.id/pressrelease/2023/02/06/1997/ekonomi-indonesia-tahun-2022-tumbuh-5-31-persen.html

Bank Indonesia. (2023). Laporan Perekonomian Provinsi. Www.bi.go.id. https://www.bi.go.id/id/publikasi/laporan/lpp/default.aspx

Chen, Z., Hu, L., Lin, R., & Chen, Z. (2021). The industrial structure upgrading and regional economic growth: An empirical analysis from china's provincial panel data. Paper presented at the 181-184. https://doi.org/10.1109/BDEIM55082.2021.00043

Florida, R., Mellander, C., & Stolarick, K. (2008). Inside the black box of regional development—human capital, the creative class and tolerance. Journal of Economic Geography, 8(5), 615-649. https://doi.org/10.1093/jeg/lbn023

Gafur , A., & Rochaida, E. (2022). Determinan tingkat pertumbuhan ekonomi dan tingkat kemiskinan. FORUM EKONOMI, 24(1), 92–102. https://doi.org/10.30872/jfor.v24i1.10393

Gennaioli, N., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2013). HUMAN CAPITAL AND REGIONAL DEVELOPMENT. The Quarterly Journal of Economics, 128(1), 105–164. https://www.jstor.org/stable/26372495

Keynes, J. (1936): The General Theory of Employment, Interest and Money, Prometheus Books, Amherst, NY. Reprinted 1997

Kim, A., Lim, J., & Colletta, A. (2023). How regional economic structure matters in the era of COVID-19: Resilience capacity of U.S. states. The Annals of Regional Science, 70(1), 159-185. https://doi.org/10.1007/s00168-022-01134-w

Media, K. C. (2022, December 29). Menyelesaikan Tantangan Geopolitik 2023 dengan “Cara Asia” ala Indonesia. KOMPAS.com. https://nasional.kompas.com/read/2022/12/29/08000051/menyelesaikan-tantangan-geopolitik-2023-dengan-cara-asia-ala-indonesia

Robert M. Solow, A Contribution to the Theory of Economic Growth, The

Quarterly Journal of Economics, Volume 70, Issue 1, February 1956, Pages 65–94, https://doi.org/10.2307/1884513
Rodríguez-Pose, A. (2013). Do institutions matter for regional development? Regional Studies, 47(7), 1034-1047. https://doi.org/10.1080/00343404.2012.748978
Romer, P. M. (1990). Endogenous technological change. Journal of political Economy, 98(5, Part 2), S71-S102.
Sari, R. D. P., & Oktora, S. I. (2021). Determinan Produktivitas Tenaga Kerja Industri Manufaktur Besar dan Sedang di Pulau Jawa. Jurnal Ekonomi Dan Pembangunan Indonesia, 21(2), 4.