Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

Application Biplot and K-Medians Clustering to Group Export Destination Countries in Asia, Africa, America, and Europe of Indonesia's Product

Rahmi Lathifah Islami[1], Pardomuan Robinson Sihombing[2]

[1]Universitas Padjadjaran

[2*]Badan Pusat Statistik, Jakarta

*robinson@bps.go.id

Abstract. A good increasing export will yield foreign exchange to a country, and subsequently funding its country growth. In Indonesia, export is one of the biggest foreign contributors. As we can see that the countries Indonesia export to are more than 100, it is a must to group them based on their similarity. Biplot and cluster analysis are statistic methods which are used as tool to classify data based on variable explanatory. There are outliers in data acquired. Outliers are observation data which is appeared to be extremely different to the other data. Those data are identified by leverage method. in summary, this research applies K-Medians Clustering Method using Manhattan Distance to resolve outliers while grouping the countries based on their export data. The data contains export data of 182 countries in the year of 2017. R 3.5.1 software was used to calculate in this analysis. The clustering shows us that each continent form difference clusters. Asia has 4 clusters while the rest each has 3 clusters. In addition, we can conclude that several clusters have high value export of Indonesia for certain variables.

Keywords: Biplot, Clustering, K-medians, Silhouette Coefficient, Export

(Received 2021-04-06, Accepted 2021-04-30, Available Online by 2021-04-30)

1. Introduction

Business Development data is really important to a country. It will affect country's annual income or economic growth. One of these business development data is country's export data. A good increasing export will yield foreign exchange to a country, and subsequently funding its country growth. In Indonesia, export is one of the biggest foreign contributors. Based on Badan Pusat Statistik [1], cumulatively Indonesia export value from January to November 2017 reached US$153.90 Billion. It is 17.16% increase year on year (*y-on-y*). Meanwhile, in January-November 2018 its value reached US$150.88 Billion or 8.84% increase in yoy. As we can see that the countries Indonesia export to are more than 100, it is a must to group them based on their similarity. Furthermore, the groups can be used

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

as consideration to take decision and policy as Indonesia keep export goods into those countries.

Biplot and cluster analysys are statistic methods which are used as tool to classify data based on variable explanatory. Biplot analysys purpose is to group data based upon their similarity which have been mapped on 2 dimensions diagram. so, biplot analysys is easier to be understood and analyzed compared to other classification method. Meanwhile clustering analysys purpose is to identify a group of objects which have certain character similarity which can be separated from the other group of objects. It caused the object in the same group tend to be homogen compared to other group of objects. The objects of this research are export activities of countries in 4 continents: Asia, Africa, America, and Europe.

There are outliers in data acquired. Outliers are observation data which is appeared to be extremely different to the other data. Those data are identified by leverage method. in summary, this research applies K-Medians Clustering Method using Manhattan Distance to resolve outliers while grouping the countries based on their export data.

2. Methods

2.1 Research Variables

The Data that is used in this research is secondary data obtained from http://trademap.org which only the export data was being taken. The data contains export data of 182 countries which are consisted of 48 Asia countries, 52 Africa countries, 42 America countries and 40 Europe countries in the year of 2017. There are 15 variables which are the most export in that data:

Table 1. Research Variable

X1: Animal or vegetable fats and oils and their cleavage products; prepared edible fats; animal or vegetable waxes (code 15)
X2: Ores, slag and ash (code 26)
X3: Mineral fuels, mineral oils and products of their distillation; bituminous substances; mineral wax (code 27)
X4: Miscellaneous chemical products (code 38)
X5: Rubber and articles thereof (code 40)
X6: Wood and articles of wood; wood charcoal (code 44)
X7: Paper and paperboard; articles of paper pulp, of paper or of paperboard (code 4
X8: Articles of apparel and clothing accessories, knitted or crocheted (code 61)
X9: Articles of apparel and clothing accessories, not knitted or crocheted (code 62)
X10: Footwear, gaiters and the like; parts of such articles (code 64)
X11: Natural or cultured pearls, precious or semi-precious stones, precious metals, metals clad with precious metal, and articles thereof; imitation jewellery; coin (code71)
X12: Iron and steel (code72)
X13: Machinery, mechanical appliances, nuclear reactors, boilers; parts thereof (code 84)
X14: Electrical machinery and equipment and parts thereof; sound recorders and reproducers, television image and sound recorders and reproducers, and parts and accessories of such articles (code 85)
X15: Vehicles other than railway or tramway rolling stock, and parts and accessories (code 87)

2.2. Biplot Analysis

Biplot analysis is introduced for the first time by [2] which has been developed with Principle Component Analysis (PCA). The biplot display is a graph of row and column markers obtained from data that forms a two-way table. The markers are calculated from the singular value decomposition of the data matrix.

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

2.3 K-medians analysis

In cluster analysis there is a need of distance type which is suitable to the analysis method so there will be optimal analysis results. Based on [3], it says the clustering using Manhattan Distance results more optimal if applied to k-medians clustering technique, because the distortion or deviation resulted is smaller than other distance. The same report was found in [4]. They found the use of median to measure center point is same as objective function of manhattan distance.
K-medians algorithm needs 3 components:

1. Defined the number of K cluster

The number of K must be identified firstly because K-medians is one of non-hierarchy method. There are several ways to define the number of cluster and find the quality and the power of a clustering, one of those ways is Silhouette Coefficient. Silhouette Coefficient is a combination of 2 methods which are cohesion method and separation method. Cohesion method is used to measure how close the relation of objects in a cluster, meanwhile separation method is used to measure how far a cluster separated from the others is [5]. The steps to measure Silhouette Coefficient is shown in the following:

a) Calculate the average distance between objects within the cluster:

$$a(i) = \frac{1}{[A]-1} \sum j \in_{A, j \neq 1} d(i,j) \quad (1)$$

Where,
a(i): the average distance between cluster members
i: an object in A cluster
j: others object in A cluster
d (i, j): distance between *i* and *j*

b) Calculate the average distance between objects within A and objects within other cluster (C), then define the minimum value:

$$d(i, C) = \frac{1}{[A]} \sum j \in C \, d(i,j) \quad (2)$$

Where,
d(i,C): the average distance between I and all objects within other cluster (C), where A≠ C

c) Calculate Silhouette Coefficient:

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \quad (3)$$

Where,
b(i): the smallest distance between cluster member with the closest cluster member. Subjective criteria of clustering quality based on silhouette coefficient that was made by [6] is shown in table 2.1.

Table 2. Subjective criteria of clustering quality based on silhouette coefficient

Silhouette Coefficient	Kauffman Interpretation
0.71 – 1.00	Strong
0.51 – 0.70	Good
0.26 – 0.50	Weak
0 – 0.25	Bad

2. Define centroid

Based on Teknomo [7], centroid could be picked randomly from all of observation.

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

3. Define the distance of each observation unit to each centroid

Distance calculation is used to locate the observation into the cluster based on the nearest centroid. Distance calculation which is being used in k-medians method is manhattan [8]. The Manhattan Distance is a distance of a point to the other point in the Cartesian. Following the vertical and horizontal axes, with no comeback. As instance, if known 2 points, which are p_1 and p_2 in two dimensions' plane (x_1, y_1) and (x_2, y_2), then the manhattan distance between p_1 and p_2 is $|x_1 - x_2| + |y_1 - y_2|$. Each distance observation is calculated to each centroid using Manhattan Distance with this following equation

$$d(\vec{x}, \vec{y}) = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_p - y_p| \quad (4)$$

Where,

$d(\vec{x}, \vec{y})$: the distance between data and each centroid

x : export proportion data

y : *centroid*

p : number of variables

Each observation is grouped based on the closest distance between the data and the centroid. Update the centroid of each cluster using median calculation from observations in the same centroid. Repeat the steps until no member cluster movement left.

3. Results and Discussion

3.1 Asia Continent

This research uses 48 countries in Asia Continent. The average value and the standard deviation from each variable is calculated before clustering the countries of Asia using k-medians clustering. The Calculation is shown in Table 3.1. as we can see in the table, there is little difference of average value and standard deviation which shown the variance in the data. X_{14} is the most popular variable in Asia meanwhile X_2 is the least. The highest variance was found in X_1 meanwhile the least is X_8. The higher variance the more distributed the variable.

Based on R, the outliers were detected as the leverage method was used. The outliers were found in Georgia, North Korea, Nepal, State of Palestine, Philippine, and Timor Leste. Further treatment is needed to cluster these countries. The biplot analysis of Asia is shown in Figure 1. Figure 1 is showing biplot graph of Asia countries. In the graph we can see the similarity relation between countries and similarity of countries to variables. Furthermore, cluster analysis is conducted so the data in the biplot analysis would not be overlapped.

The first step to do cluster analysis is to define the number of clusters. When the first k cluster is formed, calculate at once the silhouette coefficient until the good criteria coefficient is obtained. The coefficient is calculated using R 3.5.1 software and is shown in Table 3. this number is showing how well the clustering process and the cluster quality that is formed. As we can see in the Table 4, the highest value belongs to K = 4. Therefore, the writer chose 4 clusters as the number of clusters in this continent. The value of silhouette coefficient, 0.060142649 is categorized as 'Good' based on Kauffman.

After cluster analysis was conducted using k-medians through R 3.5.1 software, we can see the names of countries in Asia for each of 4 clusters the Table 5. To measure the classification of sum of export proportion in Asia for each cluster, the following can be interpretated:

a. Good Criteria if the value is far above the average of overall, and marked as '++'
b. '+' is given for the value below the '++' as long as it is still above overall average.
c. For the value under the average of overall, is given '-'mark
d. And for the value that is far under the overall average is given '—' mark

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

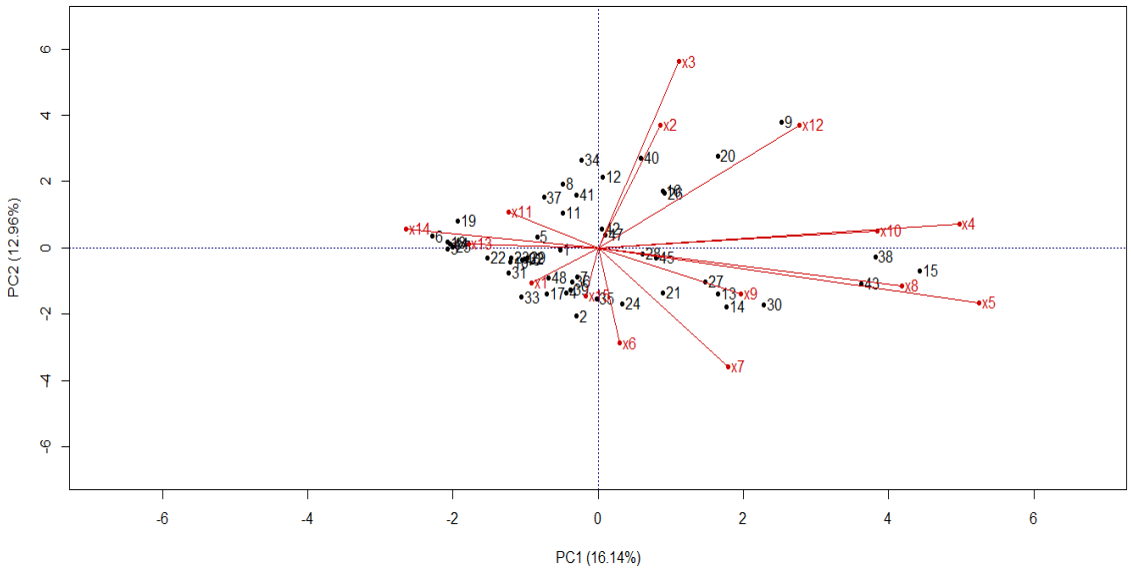

Figure 1 Biplot graph of Asiacountries

Table 3 Variable description of Asia

Variable	Standard Deviation	Mean
X_1	0.267767	0.22378
X_2	0.028901	0.008757
X_3	0.207446	0.119571
X_4	0.029024	0.018748
X_5	0.06051	0.04949
X_6	0.187337	0.092494
X_7	0.106071	0.08379
X_8	0.014934	0.009312
X_9	0.072084	0.020405
X_{10}	0.023495	0.011209
X_{11}	0.15334	0.037698
X_{12}	0.019599	0.009542
X_{13}	0.163303	0.875247
X_{14}	0.208736	0.94454
X_{15}	0.183955	0.705525

Tabel 4 Silhouette coefficient value of Asia

K	Silhouette Coefficient
3	0.1346155
4	0.60142649
5	0.10396336

Table 6 is showing us the level of export proportion classification between the clusters. As we can see in the Table 3.4, the characteristic of export proportion data of Asia is a focus of the main cluster which is cluster 2. Meanwhile, the rest 3 clusters tend to far from the average. In addition, there are two variables that is undesirable in the cluster 3 and 4, they are X_2 and X_3.

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

Tabel 5. Cluster Analysis of Asia Countries based Export Proportion using k-medians clustering

Cluster	Number of Countries	Name of Country
1	14	Bahrain, Brunei Darussalam, Kuwait, Laos, Lebanon, Mongolia, Oman, Philippines, Qatar, Saudi Arabia, Thailand, Timor-Leste, Vietnam
2	26	Armenia, Azerbaijan, Bangladesh, Cambodia, China, Georgia, Hong Kong, India, Iran, Israel, Japan, North Korea, SouthKorea, Kyrgyzstan, Malaysia, Myanmar, Nepal, Pakistan, Singapore, Sri Lanka, Syria, Chinese Taipei, Turkey, United Arab Emirates, Uzbekistan, Yemen
3	3	Jordan, Maldives, State of Palestine
4	5	Afghanistan, Bhutan, Kazakhstan, Macao China, Turkmenistan

In summary, the main export countries of Indonesia is the countries in cluster 2 with the main variables of X_1, X_3, X_4, X_5, X_8, X_9, X_{10}, X_{11} dan X_{12}. The second export countries of Indonesia is cluster 1 with main variables of X_2, X_7, X_{13}, dan X_{15}. The rest are the third and fourth export countries, which are cluster 3 and 4, with the main carriable is X_6 and X_{14}, respectively.

Table 6. The level of export proportion classification between the clusters of Asia

Variable	Mean	Cluster				Classification			
		C1	C2	C3	C4	C1	C2	C3	C4
X_1	0.22378	0.10020	0.30346	0.11449	0.18037	--	++	--	-
X_2	0.00875	0.01296	0.00932	0	0	++	+	--	--
X_3	0.11957	0.07227	0.17777	0	0	-	++	--	--
X_4	0.01874	0.00729	0.02560	0.02294	0.00899	--	++	+	-
X_5	0.04949	0.04347	0.06083	0.02794	0.01679	-	++	--	--
X_6	0.09249	0.05842	0.06088	0.66796	0.00649	-	-	++	--
X_7	0.08379	0.10550	0.08836	0.06686	0.01277	++	+	-	--
X_8	0.00931	0.00918	0.01127	0.00653	0.00072	-	++	--	--
X_9	0.02040	0.00971	0.02823	0.01784	0.00743	--	++	-	--
X_{10}	0.01120	0.01164	0.01325	0.00831	0.00078	+	++	-	--
X_{11}	0.03769	0.00070	0.06651	0.00146	0	--	++	-	--
X_{12}	0.00954	0.00953	0.01206	0.00277	0	-	++	--	--
X_{13}	0.07595	0.09313	0.0748	0.01965	0.07126	++	-	--	-
X_{14}	0.11115	0.06646	0.0405	0.00957	0.66960	-	--	--	++
X_{15}	0.128103	0.399476	0.02708	0.033608	0.02475	++	--	-	--

3.2 Africa Continent

This research uses 52 countries in the Africa Continent. The average value and standard deviation of each variable are calculated before the clustering (with k-medians clustering). The results are shown in Table 7. The highest mean is found on X_1 as it is the most popular variable in Africa, meanwhile the least is shown by X_2. The distribution of variables in Africa is the most varied compared to 3 other continents. X_1 shows the most varied variables among the others, meanwhile X_2 is the least. Based on R, outliers were found using leverage method. They are Botswana, Burkina Faso, Cabo Verde, Egypt, Malawi, Mali, Mauritania, Seychelles, Somalia, and South Africa. They need further treatment to be clustered using k-medians.

The biplot analysis of Africa was shown in Figure 2. The graph shows us the relation of similarity of each countries and similarity of countries with variables. The cluster analysis was conducted to map the data clearly as the data in biplot analysis tend to be overlapped one to another.

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

Table 7 Africa Continent research variable description

Variable	Standard Deviation	Mean
X_1	0.311546	0.6525
X_2	5.67E-05	7.86E-06
X_3	0.005032	0.00111
X_4	0.080258	0.023741
X_5	0.027638	0.014147
X_6	0.070435	0.020112
X_7	0.126892	0.087727
X_8	0.064997	0.01627
X_9	0.024258	0.012914
X_{10}	0.026484	0.009541
X_{11}	0.007522	0.001475
X_{12}	0.037929	0.006506
X_{13}	0.066613	0.04142
X_{14}	0.22428	0.087251
X_{15}	0.06685	0.025312

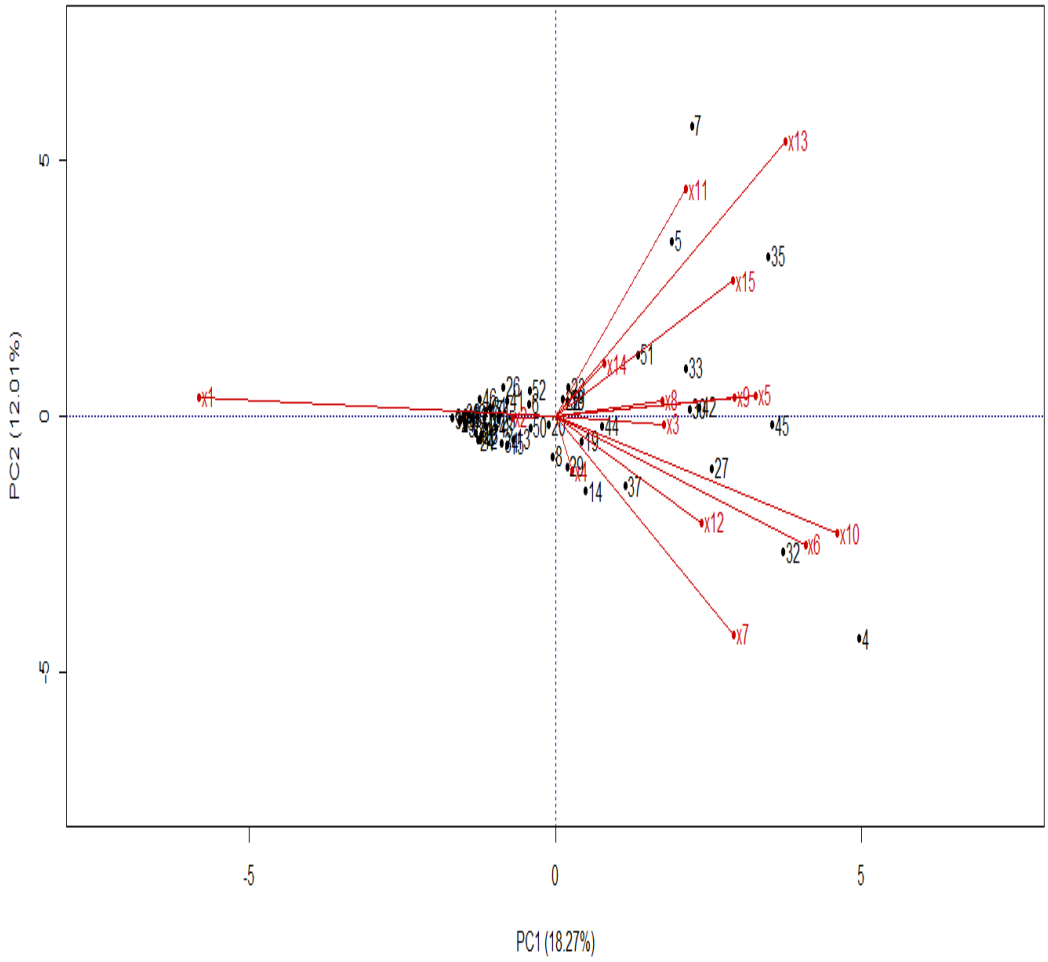

Figure 2. Biplot analysis graph of Africa

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

The first step to do cluster analysis is to define the number of clusters. When the first k cluster was formed, calculate at once the silhouette coefficient until the good criteria coefficient was found. Silhouette coefficients were obtained using R 3.5.1 and were shown in Table 8.

Table 8. Silhouette coefficient value of Africa

K	Silhouette Coefficient
3	0.88830048
4	0.1059707
5	0.3543382

As we can see in the Table 8, the highest value belongs to K = 3. Therefore, the writer chose 3 clusters as the number of clusters in this continent. The value of silhouette coefficient, 0.88830048 is categorized as 'Strong' based on Kauffman. After cluster analysis was conducted using k-medians through R 3.5.1 software, we can see the names of countries in Africa for each of 3 clusters the Table 9.

To measure the classification of sum of export proportion in Africa for each cluster, the following can be interpretated:

a. If the value is above the average of overall, and marked as '+'
b. For the value under the average of overall, is given '-' mark
c. And for the value that is far under the overall average is given '—' mark

The level of export proportion classification between the clusters of Africa was shown in Table 3.8. As we can see from the table, the characteristic of export proportion data in Africa focused on the main cluster, which is cluster 2. The cluster 2 is above the average, meanwhile the rest show the contrary. In addition, we can find a lot of variables in cluster 3 that are least desired. They are X_1, X_2, X_3, X_6, X_7, X_9, X_{11}, X_{12} dan X_{15}.

Table 9 Cluster Analysis of Africa Countries based Export Proportion using k-medians clustering

Cluster	Number of Countries	Name of Countries
1	32	Algeria, Angola, Benin, Burundi, Central African Republic, Chad, Comoros, Congo, Democratic Republic of the Congo, Djibouti, Egypt, Equatorial Guinea, Gabon, Gambia, Guinea-Bissau, Kenya, Liberia, Madagascar, Mauritania, Mozambique, Niger, Rwanda, Sao Tome and Principe, Senegal, Sierra Leone, Somalia, Swaziland, Tanzania, Togo, Tunisia, Uganda, Zimbabwe
2	17	Botswana, Burkina Faso, Cabo Verde, Cameroon, Côte d'Ivoire, Ethiopia, Ghana, Guinea, Libya, Malawi, Mali, Mauritius, Morocco, Namibia, Nigeria, Seychelles, South Africa
3	3	Eritrea, Saint Helena, Zambia

In summary, the main cluster of export countries of Indonesia in Africa is the cluster 2 with the main variables of X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13} dan X_{15}. Cluster 1 become the second export target with the main variables of X_1 and X_2, while the cluster 3 become the third with the main variables of X_{13} and X_{14}

Table 10 The level of export proportion classification between the clusters of Africa

Variable	Mean	Cluster			Classification		
		C1	C2	C3	C1	C2	C3
X_1	0.6525	0.8653	0.3668	0	+	-	--
X_2	7.86E-06	1.28E-05	0	0	+	--	--
X_3	0.0011	0.0007	0.0019	0	-	+	--

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

X_4	0.0237	0.0132	0.0475	0.0002	-	+	--
X_5	0.0141	0.0076	0.0288	0.0002	-	+	--
X_6	0.0201	0.0039	0.0540	0	-	+	--
X_7	0.0877	0.0455	0.1809	0.0091	-	+	--
X_8	0.0162	0.0030	0.0432	0.0042	--	+	-
X_9	0.0129	0.0095	0.0215	0	-	+	--
X_{10}	0.0095	0.0017	0.0244	0.0085	--	+	-
X_{11}	0.001475	1.79E-06	0.0045	0	-	+	--
X_{12}	0.006506	0.000923	0.018162	0	-	+	--
X_{13}	0.0414	0.0163	0.0867	0.052288	-	+	+
X_{14}	0.0872	0.0277	0.0513	0.9252	--	-	+
X_{15}	0.0253	0.0040	0.0697	0	-	+	--

.

3.3 America Continent

This research uses 42 countries in the America Continent. The average value and standard deviation of each variable are calculated before the clustering (with k-medians clustering). The results are shown in Table 11. The highest mean is found on X_{15} as it is the most popular variable in America, meanwhile the least is shown by X_2. The most varied variables in America was found in X_1, meanwhile X_2 is the least. Based on R, outliers were found using leverage method. They are British Virgin Islands, Canada, Cuba, Mexico, Saint Lucia, and United States of America. They need further treatment to be clustered using k-medians.

The biplot analysis of America was shown in Figure 3. The graph shows us the relation of similarity of each countries and similarity of countries with variables. The cluster analysis was conducted to map the data clearly as the data in biplot analysis tend to be overlapped one to another.

Tabel 11 America Continent research variable description

Variable	Standard Deviation	Mean
X_1	0.232036	0.133325
X_2	0.000232	3.68E-05
X_3	0.011614	0.002628
X_4	0.014868	0.010278
X_5	0.212741	0.11665
X_6	0.260717	0.11811
X_7	0.189049	0.103628
X_8	0.05618	0.023826
X_9	0.040393	0.023006
X_{10}	0.139164	0.074092
X_{11}	0.007889	0.00259
X_{12}	0.009216	0.003147
X_{13}	0.069327	0.05992
X_{14}	0.108067	0.07306
X_{15}	0.321278	0.2557

The first step to do cluster analysis is to define the number of clusters. When the first k cluster was formed, calculate at once the silhouette coefficient until the good criteria coefficient was found. Silhouette coefficients were obtained using R 3.5.1 and were shown in

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

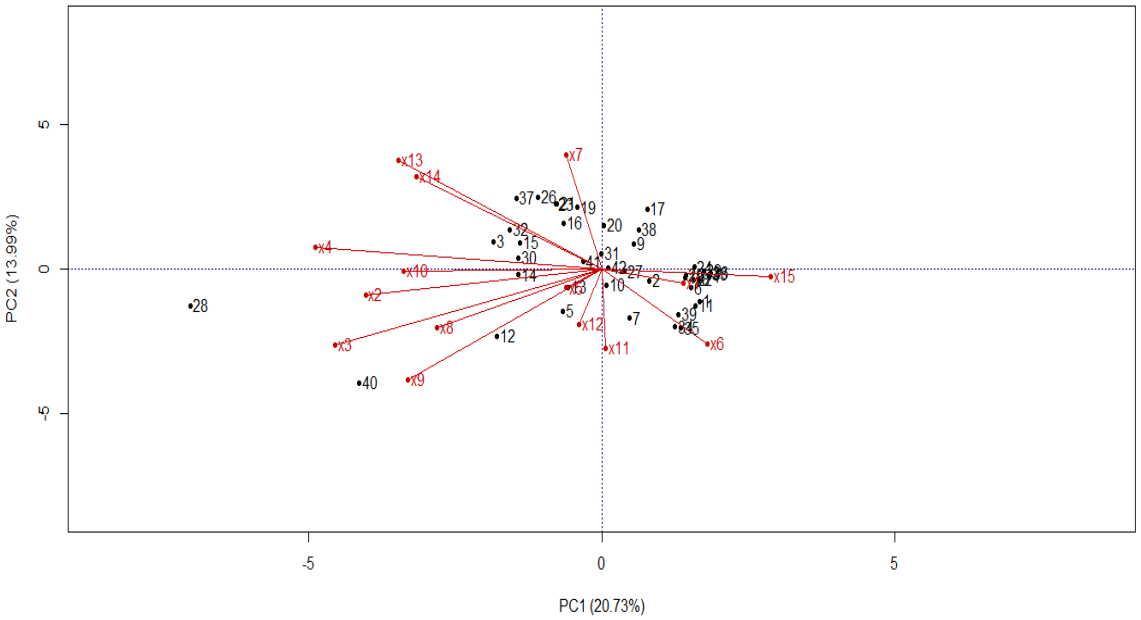

Figure 3. Biplot Analysis graph of America

Table 12. Silhouette coefficient value of America

K	Silhouette Coefficient
3	0.7140413
4	0.34643963
5	0.3640942

As we can see in the Table 3.10, the highest value belongs to K = 3. Therefore, the writer chose 3 clusters as the number of clusters in this continent. The value of silhouette coefficient, 0.7140413 is categorized as 'Strong' based on Kauffman. After cluster analysis was conducted using k-medians through R 3.5.1 software, we can see the names of countries in America for each of 3 clusters the Table 3.11.

To measure the classification of sum of export proportion in America for each cluster, the following can be interpretated:

a. If the value is above the average of overall, and marked as '+'
b. For the value under the average of overall, is given '-' mark
c. And for the value that is far under the overall average is given '—' mark

The level of export proportion classification between the clusters of America were shown in Table 3.12. As we can see from the table, the characteristic of export proportion data in America focused on the main cluster, which is cluster 2. The variables of cluster 2 almost all are above the average. Meanwhile the rest, cluster 1 and 3, show the contrary. In addition, we can find a lot of variables in cluster 1 and 3 that are least desired. They are X_2, X_3, X_{13}, X_{14} and X_2, X_3, X_{12}, respectively.

In summary, the main cluster of export countries of Indonesia in America is the cluster 2 with the main variables of X_1, X_2, X_3, X_4, X_5, X_7, X_8, X_9, X_{10}, X_{13} dan X_{14}. Cluster 1 become the second export target with the main variables of X_6, X_9 dan X_{12}. while the cluster 3 become the third with the main variables of X_{11} dan X_{15}.

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

Table 13 Cluster analysis of America countries based Export Proportion using k-medians clustering

Cluster	Number of Countries	Name of Countries
1	6	Anguilla, Bahamas, Barbados, British Virgin Islands, Saint Vincent and the Grenadines, Turks and Caicos Islands
2	28	Antigua and Barbuda,Argentina, Belize, Brazil, Canada,Cayman Islands, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, El Salvador, Grenada, Guatemala, Haiti, Honduras, Jamaica, Mexico, Panama, Paraguay, Peru, Suriname, Trinidad and Tobago , United States of America , Uruguay, Venezuela
3	8	Aruba, Bermuda, Bolivia, Guyana, Nicaragua, Saint Kitts and Nevis, Saint Lucia, St Pierre and Miquelon

Table 14. The level of export proportion classification between the clusters of America

Variable	Mean	Cluster			Classification		
		C1	C2	C3	C1	C2	C3
X_1	0.1333	0.0816	0.1664	0.056	-	+	--
X_2	3.68E-05	0	5.51E-05	0	--	+	--
X_3	0.0026	0	0.0039	0	--	+	--
X_4	0.0102	0.0063	0.0137	0.001	-	+	--
X_5	0.1166	0.0056	0.1709	0.009	--	+	-
X_6	0.1181	0.7017	0.0206	0.021	+	-	-
X_7	0.1036	0.0049	0.1520	0.008	--	+	-
X_8	0.0238	0.0024	0.0350	0.0004	-	+	--
X_9	0.0230	0.0280	0.0264	0.007	+	+	-
X_{10}	0.0740	0.0332	0.1037	0.0009	-	+	--
X_{11}	0.0025	0.0021	0.0018	0.005	-	--	+
X_{12}	0.0031	0.0076	0.0030	0	+	-	--
X_{13}	0.0599	0	0.0889	0.003	--	+	-
X_{14}	0.0730	0	0.0958	0.048	--	+	-
X_{15}	0.2557	0.125	0.1171	0.837	-	-	+

3.4 Europe Continent

This research uses 40 countries in the Europe Continent. The average value and standard deviation of each variable are calculated before the clustering (with k-medians clustering). The results are shown in Table 15.

The highest mean is found on X_1 as it is the most popular variable in Europe, meanwhile the least is shown by X_2. The most varied variables in Europe was found in X_5, meanwhile X_2 is the least. Based on R, outliers were found using leverage method. They are Belarus, Ireland, Slovenia, Switzerland, and Ukraine. They need further treatment to be clustered using k-medians.

The biplot analysis of Europe was shown in Figure 4. The graph shows us the relation of similarity of each countries and similarity of countries with variables. The cluster analysis was conducted to map the data clearly as the data in biplot analysis tend to be overlapped one to another.

Table 15. Europe Continent research variable description

Variable	Standard Deviation	Mean
X_1	0.259281	0.200674
X_2	0.002923	0.000479
X_3	0.042051	0.01128
X_4	0.046908	0.02561

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

X_5	0.251331	0.20975
X_6	0.038053	0.025969
X_7	0.135202	0.07002
X_8	0.05489	0.04398
X_9	0.053991	0.038759
X_{10}	0.161764	0.117964
X_{11}	0.154227	0.027231
X_{12}	0.00647	0.003387
X_{13}	0.111575	0.05473
X_{14}	0.190139	0.138673
X_{15}	0.069105	0.03149

The first step to do cluster analysis is to define the number of clusters. When the first k cluster was formed, calculate at once the silhouette coefficient until the good criteria coefficient was found. Silhouette coefficients were obtained using R 3.5.1 and were shown in Table 16.

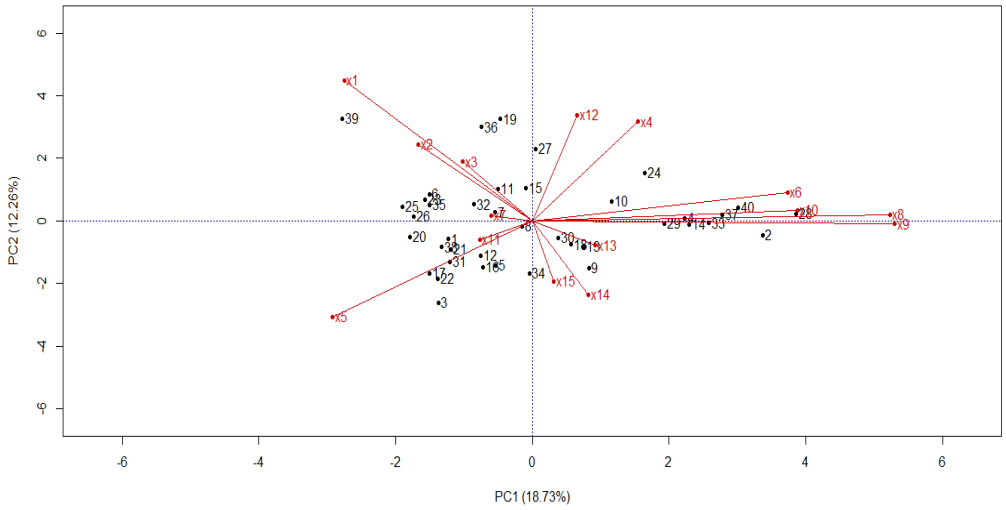

Figure 3 Biplot Analysis graph of Europe

Table 16. Silhouette coefficient value of Europe

K	Silhouette Coefficient
3	0.3688017
4	0.3599565
5	0.1291649

As we can see in the Table 17, the highest value is belong to K = 3. Therefore, the writer chose 3 clusters as the number of clusters in this continent. The value of silhouette coefficient, 0.3688017 is categorized as 'Weak' based on Kauffman. After cluster analysis was conducted using k-medians through R 3.5.1 software, we can see the names of countries in Europe for each of 3 clusters the Table 3.15.

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

Tabel 17 Cluster analysis of europe countries based Export Proportion using k-medians clustering

Cluster	Number of Countries	Name of Countries
1	12	Albania, Croatia, Estonia, Greece, Italy, Macedonia, Moldova, Montenegro, Netherlands, Russian Federation, Spain, Ukraine
2	16	Austria, Belgium, Cyprus, Czech Republic, Denmark, France, Germany , Ireland, Malta, Norway, Poland, Serbia, Slovakia, Sweden, Switzerland, United Kingdom
3	12	Belarus, Bosnia and Herzegovina, Bulgaria, Finland , Hungary, Iceland , Latvia, Lithuania, Luxembourg, Portugal, Romania, Slovenia

To measure the classification of sum of export proportion in Europe for each cluster, the following can be interpretated:

a. If the value is above the average of overall, and marked as '+'
b. For the value under the average of overall, is given '-' mark
c. And for the value that is far under the overall average is given '—' mark

The level of export proportion classification between the clusters of Europe were shown in Table 3.16. As we can see from the table, the characteristic of export proportion data in America focused on the main cluster, which is cluster 2. The variables of cluster 2 tend to be above the average. Meanwhile the rest, cluster 1 and 3, show the contrary. In addition, we can find a variable in cluster 3 which is undesired. The variable is X_2.

In summary, the main cluster of export countries of Indonesia in Europe is the cluster 2 with the main variables of X_4, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{13}, X_{14}, dan X_{15}. Cluster 1 become the second export target with the main variables of X_1, X_2, X_3, X_4, X_7 dan X_{12}, while the cluster 3 become the third with the main variables of X_3, X_5 dan X_{15}.

Table 18 The level of export proportion classification between the clusters of Europe

Variable	Mean	Cluster			Classification		
		C1	C2	C3	C1	C2	C3
X_1	0.200674	0.536578	0.045614	0.071516	+	--	-
X_2	0.0004	0.0015	2.9E-05	0	+	-	--
X_3	0.01128	0.015845	0.001177	0.020197	+	-	+
X_4	0.0256	0.0342	0.0315	0.0090	+	+	--
X_5	0.2097	0.0877	0.0523	0.5416	--	--	+
X_6	0.0259	0.0185	0.0438	0.0095	-	+	--
X_7	0.0700	0.0946	0.0702	0.0450	+	+	-
X_8	0.0439	0.0118	0.0884	0.0167	-	+	-
X_9	0.0387	0.0156	0.0814	0.0049	-	+	--
X_{10}	0.1179	0.0379	0.2033	0.0841	--	+	-
X_{11}	0.0272	0.0009	0.0666	0.0009	--	+	--
X_{12}	0.0033	0.0051	0.0030	0.002	+	-	--
X_{13}	0.0547	0.0375	0.0837	0.0332	--	+	--
X_{14}	0.1386	0.0874	0.1931	0.1172	--	+	-
X_{15}	0.0314	0.0141	0.0351	0.0438	-	+	+

4. Conclusion

The clustering shows us that each continent form difference clusters. Asia has 4 clusters while the rest each has 3 clusters. In addition, we can conclude that several clusters have high value export of Indonesia for certain variables. So, these research could be utilized as consideration to make policies related to export.

Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 3, No.1, April 2021, pp. 0210105-01 ~ 0210105-14
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v3i1.8451

Reference

[1]. Badan Pusat Statistik.(2017).Perkembanga Ekspor Impor Indonsia.
[2]. Gabriel, K, R. (1971), "The Biplot Graphic Display of Matrices with Aplication to Principal Component Analysis," Journal of Biometrica. 58, 453-467.
[3]. Anderson, Gross, Musicant, Ritz, Smith, dan Steinberg. (2006), "Adapting K-Medians to Generate Normalized Cluster Centers."
[4]. Wierzchon and Klopotek. (2017), "Modern Algorithms of Cluster Analysis," Switzerland: Springer.
[5]. Pramesti, Dyang Falila, M. Tanzil Furqon, and Candra Dewi. (2017), "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X Vol. 1, No. 9, 723-732.
[6]. Kaufman, L., Rousseeuw, PJ. (1990), "Finding Groups in Data: An Introduction to Gerombol Analysis," New York: John Wiley.
[7]. Teknomo, Kardi. (2007), "K-Means Clustering Tutorials."
[8]. Ritz, A., Anderson B., Nelson K. (2006), "Clustering Algorithms: A New Look at Mass Spectra," Computer Science Carleton College.