Higher Education Clustering in Indonesia by using k-means and Geographical Similarity Methods

[1]Purba Daru Kusuma and [2]Nilla Rachmaningrum
[1]School of Electrical Engineering, Telkom University, Bandung, Indonesia
[2]Faculty of Electrical Engineering, Institut Teknologi Telkom Surabaya, Surabaya, Indonesia

Abstract: Diverse quality and equality in higher education has been concerned by government of Indonesia. It is because these aspects have positive correlation with the country development and competitiveness. Meanwhile, improvement policy should be supported by better perspective and mapping about the condition of the higher education in this country. Although, ministry of research and higher education of Indonesia has published statistical data, the analysis of it is very limited. Based on this problem, we use clustering method to analyze this higher education statistic data, so that, new perspective and understanding can be explored. In this research, we use two computational methods: k-means and geographical similarity, so that, the analysis can be enriched. In this research, we also compare the condition in private institution and public institution. Result shows that in some aspects, there is disparity between private institution and public institution. Meanwhile in some aspect, there is disparity between Java Region and other regions in Indonesia.

Key words: Clustering, k-means, education Indonesia, geographic similarity, computational methods, institution

INTRODUCTION

Higher education quality and equality are very important aspect in Indonesia. Higher education quality affects the human resource quality, so that, they can become the development agent and compete globally. Higher education has responsibility in supporting national economic growth, national development and flourishing societies (Storen and Tremonte, 2013). Thomson enriched the function of higher education into six fields: human capital and capacity development, methodological and liberal learning, social values, transformative education and critical approach, R&D and academic research, community service (Thomson, 2008). Meanwhile, equality in higher education should be improved too, so that, the development and economic disparity among provinces in Indonesia can be reduced.

Many higher education institutions in developing countries face common problems: under-resourced, overcrowded, under-funded, weak faculty, brain drain, low paid teachers, weak infrastructure and management and lack of flexibility (Thomson, 2008). As a developing country Indonesia faces similar problem too. That is why, government of Indonesia spends big budget to solve this problem. In 2019, government spends 41 trillion rupiah to improve the quality and equality of higher education in Indonesia (THE., 2018). This budget is distributed into several posts. The 29 trillion rupiah is allocated into management and technical support program, one trillion rupiah for institution quality improvement program, 6.3 trillion rupiah for learning and student program, 1.7 trillion rupiah for human resource improvement program, 1.8 trillion rupiah for research and development strengthening program and 299 billion rupiah for the innovation strengthening program (THE., 2018).

Externally, higher education in Indonesia still needs improvement to compete with other countries in region, at least in South-East Asia moreover: Asia (THE., 2018). In 2013, higher education ranking that measures 50 countries in Asia places Indonesia at the bottom of 50 behind India, Iran and Thailand (Anonymous, 2014). Nowadays, the finest higher education institutions in Indonesia are still cannot compete with other higher education institutions in Singapore, Malaysia or Thailand (THE., 2018). Based on Times Higher Education, National University of Singapore is in the first place in Asia (THE., 2018). University of Malaya, the finest higher education institution in Malaysia is in the 46th place in Asia (THE., 2018). Mahidol University, the finest higher education institution in Thailand is in the 97th place in Asia (THE., 2018). University of the Philippines is in the 156th place in Asia (THE., 2018). Meanwhile, Bandung Institute Technology as the finest higher education institution is not in the first 200 best higher education in Asia (THE., 2018).

Corresponding Author: Purba Daru Kusuma, School of Electrical Engineering, Telkom University, Bandung, Indonesia

Meanwhile, Indonesia is a very wide archipelagic country with 1,910,931 km^2 covering area and 17,504 islands (Moeliodihardjo, 2010). The development progress disparity among provinces is still severe (Aritenang, 2008). In his research, Aritenang noted that poor province has higher GDP per capita growth and grows faster than rich province (Aritenang, 2008). Disparity in development affects disparity in higher education quality among provinces. Many good higher education institutions are in the Island of Java. It is because the population in this Island is very high and the development level of the provinces in the Island of Java basically is higher than provinces outside the Island of Java. This condition also makes many students from outside the Island of Java enter this Island to get better higher education quality. Higher education is also blamed as one of triggers in class divide in Indonesia (Anonymous, 2014).

Besides geographical aspect inequality in higher education institution also occurs between private institution and public institution. This significant disparity occurs in cost and quality (Anonymous, 2014). Most of students in Indonesia want to achieve a place at public university (Anonymous, 2014).

Based on this condition, government of Indonesia, Especially, The Ministry of Research, Technology and Higher Education should have good mapping about higher education condition. It is because the government budget is limited, so that, mistreatment should be avoided. Basically, government has had the complete statistic of higher education condition. Unfortunately, although the information has been processed, the statistic data is still presented tabular, so that, too little knowledge that can be obtained. More advance processing should be done to obtain more knowledge from this rich statistical data.

Based on this problem and needs, we cluster this statistical data of the higher education condition, so that, higher education condition in Indonesia can be grouped based on the parameters that are used. The main purpose of this research is to give better perspective and knowledge and new findings about higher education condition in Indonesia. We use k-means method as the main clustering method. Besides k-means, we also cluster this data based on the geographic similarity, so that, the result from both method can be compared to each other.

MATERIALS AND METHODS

In this research, we develop web based clustering application by using PHP language, so that, the data from government publication can be clustered. In this application, we use k-means method as a basis method. This application obtains statistical data from MySQL database. Figure 1 shows the system diagram.
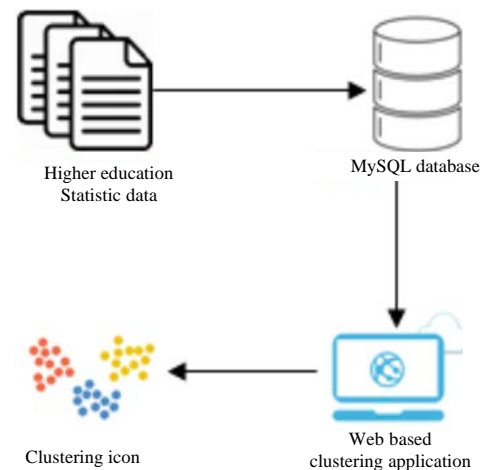

Fig. 1: System diagram

The higher education statistic data is obtained from Higher Education Statistical Year Book, 2017 (MRTHERI, 2017). This book is published by Ministry of Research, Technology and Higher Education of Indonesia. This book provides about institution, study program, new entrant, enrolled student, graduate, lecturer and drop-outs (MRTHERI, 2017). Not all of this statistical data in this book is used in this research. Data that is used in this research then is stored into MySQL database. Data is grouped based on province, so that, the clustering process is basically province clustering.

This research uses k-means method because this method is a popular quantitative clustering method, especially in processing education data (Ogwoka *et al.*, 2015; Durairaj and Vijitha, 2014; Bhogan *et al.*, 2017). Besides that, this method is also used in our previous clustering work. All of these researches aim to predict student's academic performance (Ogwoka *et al.*, 2015; Durairaj and Vijitha, 2014; Bhogan *et al.*, 2017). By combining k-means with decision tree, they can improve the accuracy and execution time rather than the previous research that used ID3 and C4.5 algorithm (Ogwoka *et al.*, 2015). Meanwhile, Durairaj and Vijitha combined k-means method with Naive Bayes or decision tree. In their research, they showed that Naive Bayes performed better result rather than decision tree (Durairaj and Vijitha, 2014). Bhogan *et al.* (2017) used combined k-strange and Naive Bayes method to improve the disadvantage of k-means-decision tree combined method (Bhogan *et al.*, 2017). Then, this clustering process was combined with classification process by using multiple linear regression technique (Bhogan *et al.*, 2017). k-means method is also chosen because of its easiness and simplicity. In this
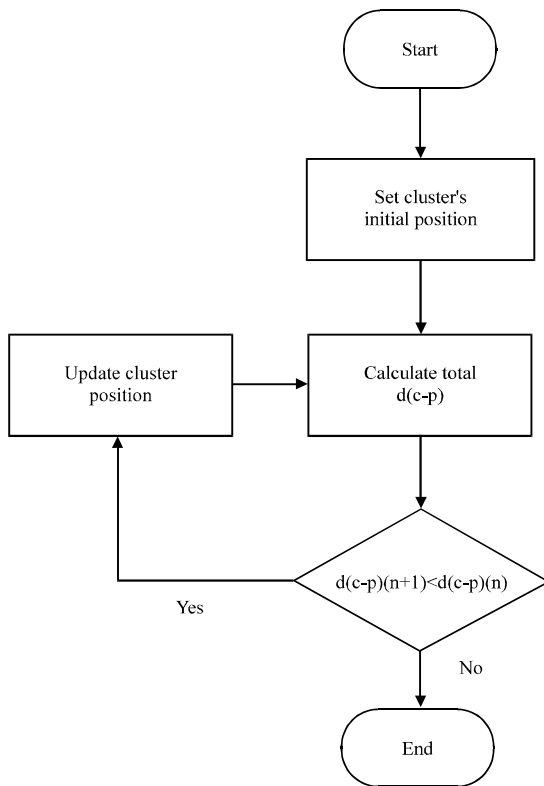
Fig. 2: k-means basic process flowchart

Table 1: Clustering process

Process	Clustered data
1	Number of public institution
2	Number of private institution
3	Number of total institution
4	Private-public institution ratio
5	Number of study program in public institution
6	Number of study program in private institution
7	Study program-institution ratio in public institution
8	Study program-institution ratio in private institution
9	Number of student in public institution
10	Number of student in private institution
11	Student-institution ratio in public institution
12	Student-institution ratio in private institution
13	Student-study program ratio in public institution
14	Student-study program ratio in private institution
15	Number of lecturers in public institution
16	Number of lecturers in private institution
17	Student-lecturer ratio in public institution
18	Student-lecturer ratio in private institution

In this research, there are twenty clustering processes. In every process, there are two clustering method: k-means based method and geographical similarity based method. The clustering process list is shown in Table 1.

As shown in Table 1, we run private institution and public institution in separate process. Many researches show that there are different characteristics between private institution and public institution, so that, the condition between both institutions tends to be different (Moeliodihardjo, 2010; Moeliodihardjo, 2014). Moeliodihardjo (2014) also classified higher education institution based on its provider: private or public (Moeliodihardjo, 2014). In public university, most of its revenue comes from government fund although in few elite public universities, government portion tends to decrease, so that, the tuition fee is as expensive as in common private institution (Moeliodihardjo, 2014). In private institution, almost the entire revenue comes from student tuition fee (Moeliodihardjo, 2014). In government accreditation based quality aspect in 2009, most of private institutions received medium or low grade (Moeliodihardjo, 2014). In the other side, most of public institutions receive medium or high grade (Moeliodihardjo, 2014).

Besides k-means method, we also use geographical similarity method, so that, clustering result from both methods can be compared to each other. Provinces in Indonesia are grouped into six clusters/groups based on region: Java, Sumatera, Kalimantan, Sulawesi, East Region and little Sunda. The members of each region are shown in Table 2 consecutively.

research, we use five clusters, so that, it represents: the lowest, low, medium, high, the highest. Figure 2 shows the general k-means clustering process.

There are several variables in this research. Variable C represents set of cluster, so that, $C = \{c_1, c_2, c_3, c_4, c_5\}$. Variables P represents set of nodes or provinces, so that, $P = \{p1, p2, p3, ..., p_m\}$. Variable m is the number of provinces in Indonesia. Variable d(c-p) is the total cluster-province distance and its value is obtained by using Eq. 1. In Eq. 1, c_{sel} is cluster where province i is in it. Variable i represents province index. Variable n is the time step:

$$d(c - p) = \sum_{i=1}^{m} \| p_i - c_{sel}(i) \| \tag{1}$$

The clustering process stops when the total distance reaches the minimum value or current total distance is not less than previous total distance. Else, the cluster should be repositioned. The cluster repositioning process is done by using Eq. 2. In Eq. 2, the next cluster j position is the average of provinces position where these provinces are in cluster j:

$$c_j(n+1) = \frac{\sum p}{n_{p,j}}, \, | \, c_{sel}(i) = c_j \tag{2}$$

Table 2: Geographical similarity group

Cluster	Cluster member
1	DKI Jakarta, West Java, Banten, Central Java, DI Yogyakarta, East Java
2	Aceh, North Sumatera, West Sumatera, Riau, Riau Island, Jambi, South Sumatera, Bangka Belitung, Bengkulu, Lampung
3	West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan
4	North Sulawesi, Gorontalo, Central Sulawesi, South Sulawesi, West Sulawesi, South-East Sulawesi
5	Maluku, North Maluku, Papua, West Papua
6	Bali, West Nusatenggara, East Nusatenggara

Table 3: k-means based clustering of public institution

Cluster	Number of member	Average score	Provinces list
1	10	1	Banten, Jambi, Bengkulu, Central Kalimantan, North Kalimantan, Gorontalo, Central Sulawesi, West Sulawesi, North Maluku, West Nusatenggara
2	6	2	Riau, Riau Island, South Sumatera, Bangka Belitung, South-East Sulawesi, West Papua
3	5	3	North Sumatera, Lampung, South Kalimantan, Maluku, Papua
4	10	4.7	DKI Jakarta, DI Yogyakarta, Aceh, West Sumatera, West Kalimantan, East Kalimantan, North Sulawesi, South Sulawesi, Bali, East Nusatenggara
5	3	12.67	West Java, Central Java, East Java

Table 4: Geographical similarity based clustering of public institution

Region	Average score	Minimum score	Maximum score
1	8.00	1	17
2	2.80	1	7
3	2.80	1	5
4	2.33	1	5
5	2.25	1	3
6	3.00	1	4

RESULTS AND DISCUSSION

After the clustering process runs, the result then is collected and is analyzed. The result of every clustering process is presented in tabular with 4 columns. The columns data are cluster number, number of members in the cluster, average score of the cluster and the list of provinces that are in the cluster.

The first clustering process is the number of public higher education institutions in a province. The purpose of this process is to observe the concern of government to build higher education in provinces in Indonesia. The k-means based clustering result is shown in Table 3. Meanwhile, the geographical similarity based clustering result is shown in Table 4.

As shown in Table 3, provinces with the highest number of public higher education institutions in Indonesia are the most populated provinces in the Island of Java. Meanwhile, there are ten provinces that only have one public higher education institution. The disparity between the first cluster and the fifth cluster is very high. The average score of the fifth cluster is approximately 12 times than the average score in the first cluster. Although, all of fifth cluster members are provinces in Java, Banten which it is province in Java is in the first cluster. Although, the economic and development level of Banten is high, this province has

only one public higher education institution. Two other provinces in Java Region (DKI Jakarta and DI Yogyakarta) are in fourth cluster.

All other Java Region provinces are distributed into first cluster to fourth cluster. From first cluster to fourth cluster, there are provinces in Sumatera in every cluster. There are not any correlation between development level and the number of public institutions in the province. Although, the economic level of the East Nusatenggara is worse than West Nusatenggara, the number of public institutions in East Nusatenggara is higher than in West Nusatenggara. The next analysis is the geographical similarity based clustering result of public institution.

As shown in Table 4, in all clusters, there is at least one province that has only one public institution. Based on this minimum number, government's minimum obligation to build and provide equality in higher education among provinces has been achieved. Meanwhile by observing the maximum score, the number of public higher education institution in Java Region is significant high. This disparity is far compared with non Java Provinces. Sumatera region achieves the second best but its maximum score is still lower than half of the Java Region maximum score. The lowest maximum score is in the East Region. Based on the maximum score analysis, it shows that there is significant disparity between Java Region and East Region. This result is caused by some

Table 5: k-means based clustering of private institution

Cluster	Number of member	Average score	Provinces list
1	8	15.88	Bangka Belitung, Bengkulu, Central Kalimantan, North Kalimantan, Gorontalo, West Sulawesi, North Maluku, West Papua
2	6	35.33	Riau Island, Jambi, Central Sulawesi, South-East Sulawesi, Maluku, Papua
3	7	51.29	West Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, Bali, West Nusatenggara, East Nusatenggara
4	7	99.14	Banten, DI Yogyakarta, Aceh, West Sumatera, Riau, South Sumatera, Lampung
5	6	293.67	DKI Jakarta, West Java, Central Java, East Java, North Sumatera, South Sulawesi

Table 6: Geographical similarity based clustering of private institution

Region	Average score	Minimum score	Maximum score
1	251.33	108	380
2	83.90	16	267
3	33.40	5	50
4	61.33	13	213
5	26.25	17	42
6	55.67	52	61

Table 7: k-means based clustering of total institution

Cluster	Number of member	Average score	Provinces list
1	8	17.13	Bangka Belitung, Bengkulu, Central Kalimantan, North Kalimantan, Gorontalo, West Sulawesi, North Maluku, West Papua
2	6	37.33	Riau Island, Jambi, Central Sulawesi, South-East Sulawesi, Maluku, Papua
3	7	54.86	West Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, Bali, West Nusatenggara, East Nusatenggara
4	7	102.57	Banten, DI Yogyakarta, Aceh, West Sumatera, Riau, South Sumatera, Lampung
5	6	302.17	DKI Jakarta, West Java, Central Java, East Java, North Sumatera, South Sulawesi

factors: population density in Java Region is very high and development progress in Java Region is also significant higher than other region in Indonesia.

The second analysis is the clustering of private higher education institution. This analysis is useful to measure the contribution of private sector/foundation in developing higher education in Indonesia. The result of k-means based clustering is shown in Table 5. Meanwhile, the result of geographical similarity method is shown in Table 6.

In Table 5, it is shown that private sector contributes significantly in provinces with high development level as shown in the fifth cluster. DKI Jakarta, West Java, Central Java and East Java are the most populous provinces in Indonesia. Their economic and development level is also the highest. Meanwhile, North Sumatera and South Sulawesi are the economic power horse provinces in their region. Provinces in the first cluster are the least populous provinces in Indonesia. This condition is rationale because private higher education institution operational capability is depended on mostly from the tuition fee. So, private sector tends to run higher education institution in the region where it can get student easily with rationale amount of the tuition fee. Unfortunately, the average score disparity between the first cluster and the fifth cluster is very wide. The fifth cluster average score is approximately twenty times the first cluster average score. It can be said that the inequality in private higher education institution is higher than in public higher education institution.

As shown in Table 6, disparity between Java Region and other regions is wide enough. Based on the average score, Java is still the region with highest number of private higher education institution. Meanwhile, East region is still the region with the lowest number of private institutions. It means the inequality in private higher education institution among regions is still high. Inside the region, little Sunda Region is the region with the smallest disparity, so that, the equality in this region is high enough. Disparity between the minimum score and the maximum score in Aceh Region and Sulawesi Region is very high with the maximum score is approximately 16 higher than the minimum score. It shows that there is significant development and economic disparity among provinces inside these regions, so that, the private sector participation in developing higher education institution in these regions is very unequal.

The 3rd analysis is the total institution. This number is obtained by summing the private institution with the public institution. The k-means based clustering result is shown in Table 7. Meanwhile, the geographical similarity based clustering result is shown in Table 8.

As shown in Table 7, the result is almost similar to the result in Table 5. The number of provinces in every cluster in Table 7 is similar to the number of provinces in every cluster in Table 5. The difference is that the average score in Table 7 is a little bit higher than in Table 5. This fact shows that in number of higher education institution aspect, private sector dominates the government. Meanwhile, Table 8 shows similar condition.

Table 8: Geographical based clustering of total institution

Region	Average score	Minimum score	Maximum score
1	259.33	112	392
2	86.700	18	270
3	36.200	6	55
4	63.670	14	218
5	28.500	18	45
6	58.670	55	65

Table 9: k-means based clustering of private-public institution ratio

Cluster	Number of member	Average score	Provinces list
1	6	8.750	Bangka Belitung, West Kalimantan, East Kalimantan, North Kalimantan, Maluku, West Papua
2	15	16.48	East Java, Aceh, West Sumatera, Riau Island, Bengkulu, Central Kalimantan, South Kalimantan, North Sulawesi, Gorontalo, West Sulawesi, South-East Sulawesi, North Maluku, Papua, Bali, East Nusatenggara
3	6	31.05	West Java, Central Java, DI Yogyakarta, Riau, Lampung, Central Sulawesi
4	5	50.24	DKI Jakarta, Jambi, South Sumatera, South Sulawesi, West Nusatenggara
5	2	103.5	Banten, North Sumatera

Table 10: Geographical similarity based clustering of private-public institution ratio

Region	Average score	Minimum score	Maximum score
1	48.00	19.29	118.00
2	32.15	8.000	89.000
3	12.67	5.000	22.000
4	23.27	13.00	42.600
5	12.38	9.000	17.000
6	27.42	13.00	54.000

The 4th analysis is the private-public institution ratio. This ratio is obtained by dividing the number of private institutions with the number of public institutions. This parameter indicates how better private sector concerns in developing higher education compared with the government. Table 9 shows the k-means based clustering result. Meanwhile, Table 10 shows the geographical similarity based clustering result.

Table 9 shows that the fifth cluster contains two provinces: Banten and North Sumatera. The analysis is as follows. The average score in Banten is high because this province has only one public higher education institution. In the other side, North Sumatera has very high average score because in North Sumatera, Batak tribe and the church have very high attention in education. Meanwhile, there are many Batak people that live outside North Sumatera are successful and have high attention in developing their place of origin. Meanwhile in the third cluster and in the fourth cluster, although, private sector attention in developing higher education is high, government attention is also high.

Table 10 shows that Java Region dominates the score (average, minimum and maximum). It is also shown that private sector is very interested in developing institution in Java Region. In this process, the second place is Sumatera Region. Unfortunately, private sector interest in developing institution in Kalimantan Region and East Region is very low. Based on this condition, to improve this ratio in these regions, government should give many incentives, so that, private sector will more interested to build higher education institution in these regions.

Result in Table 9 and 10 strengthens the previous research that said that private sector plays significant role in making higher education is more accessible in Indonesia (Moeliodihardjo, 2014). Both methods show that the private institution outnumbers public institution in all regions and in all clusters.

The fifth analysis is the number of study programs in public institution. Number of study programs indicate how diverse the field of study that is provided by government in the province. The study program ranges from diploma to doctoral degree. Table 11 shows the k-means based clustering result. Meanwhile, Table 12 shows the geographical similarity based clustering result.

Table 11 shows that there are three provinces in the fifth cluster. These provinces are the biggest provinces in Java. Compared with data in Table 3, members of the fifth cluster in Table 11 are same with the members of the fifth cluster in Table 3. Besides that there are positive relation between number of study programs and number of higher education institutions. Meanwhile, the number of members in fourth cluster is only 6 provinces. Compared with data in Table 3, it is shown that public institutions in West Kalimantan, East Kalimantan, North Sulawesi, Bali and East Nusatenggara have less study programs rather than provinces that are in the fourth cluster in Table 11. The average score of the first cluster is 3.8% of the average score of the fifth cluster. It means the disparity between the cluster with the lowest average score and the cluster with the highest average score is very wide and government still should reduce this inequality.

Table 11: k-means based clustering of number of study program in public institution

Cluster	Number of member	Average score	Provinces list
1	6	30.83	Banten, Riau Island, Bangka Belitung, North Kalimantan, West Sulawesi, North Maluku
2	9	76.67	Jambi, Bengkulu, Central Kalimantan, Gorontalo, Central Sulawesi, Maluku, Papua, West Papua, West Nusatenggara
3	10	143.5	Riau, South Sumatera, Lampung, West Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, South-East Sulawesi, Bali, East Nusatenggara
4	6	320.0	DKI Jakarta, DI Yogyakarta, Aceh, North Sumatera, West Sumatera, South Sulawesi
5	3	772.0	West Java, Central Java, East Java

Table 12: Geographical similarity based clustering of number of study program in public institution

Region	Average score	Minimum score	Maximum score
1	534.83	47	911
2	140.30	25	284
3	98.000	19	155
4	127.83	18	299
5	73.250	43	105
6	128.00	61	205

Table 13: k-means based clustering of number of study program in private institution

Cluster	Number of member	Average score	Provinces list
1	5	51.6000	Bangka Belitung, Central Kalimantan, North Kalimantan, Gorontalo, West Sulawesi
2	13	131.460	Riau Island, Jambi, Bengkulu, West Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, Central Sulawesi, South-East Sulawesi, Maluku, North Maluku, Papua, West Papua
3	10	357.700	Banten, DI Yogyakarta, Aceh, West Sumatera, Riau, South Sumatera, Lampung, Bali, West Nusatenggara, East Nusatenggara
4	3	1,040.00	Central Java, North Sumatera, South Sulawesi
5	3	1,768.67	DKI Jakarta, West Java, East Java

Table 14: Geographical similarity based clustering of number of study program in private institution

Region	Average score	Minimum score	Maximum score
1	1,284.50	502	1.916
2	312.600	32	1.078
3	106.400	15	165
4	226.500	58	766
5	114.000	100	143
6	263.330	250	277

Table 12 shows that Java Region provides the highest number of study programs in public institutions compared with other regions. East Region is region with the least average score. But, the minimum score in Java Region is smaller than the minimum score in little Sunda Region. In the ratio between the minimum score and the maximum score, East Region gets the highest ratio. Meanwhile, Java gets the lowest ratio. It means that in number of study programs in public higher education institution aspect, East Region is the most equal region. Meanwhile, Java is the least equal one.

The 7th analysis is the number of study programs in private institution. The number of study programs indicates how diverse the field of study that is provided by private sector in related province. The study program ranges from diploma to doctoral degree. Table 13 shows the k-means based clustering result. Meanwhile, Table 14 shows the geographical similarity based clustering result.

Table 13 shows that the study program distribution in private institution is just little bit different to the distribution in public institution. Provinces that are in the fifth cluster are same. But, the disparity is much wider. Although, the average score of the fourth cluster is only 58 % of the average score of the fifth cluster, the average score score of the first cluster is 2.8% of the average score of the fifth cluster. It means that the disparity between the cluster with the lowest average score and the cluster with the highest average score is wider than that occurs in public higher education one. It also shows that private sector view is more practical rather than the government view. To reduce this inequality, government should give more incentives to private sector that has intention to open new study program, especially in provinces where the number of study programs is still low.

Table 14 shows that Java Region still achieves the first place in number of study programs aspect. The second place belongs to Sumatera Region. Although, the

Table 15: k-means based clustering of study program-institution ratio in public institution

Cluster	Number of member	Average score	Provinces list
1	4	16.50	Riau Island, Bangka Belitung, North Kalimantan, West Sulawesi
2	10	35.07	Aceh, West Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, Maluku, North Maluku, Papua, West Papua, East Nusatenggara
3	10	54.54	Banten, East Java, West Sumatera, Riau, Lampung, Central Kalimantan, South Sulawesi, South-East Sulawesi, Bali, West Nusatenggara
4	5	70.77	West Java, Central Java, South Sumatera, Bengkulu, Central Sulawesi
5	5	89.76	DKI Jakarta, DI Yogyakarta, North Sumatera, Jambi, Gorontalo

Table 16: Geographical similarity based clustering of study program-institution ratio in public institution

Region	Average score	Minimum score	Maximum score
1	70.76	47.00	109.75
2	54.30	12.50	87.000
3	35.63	19.00	52.000
4	56.97	18.00	90.000
5	33.88	27.50	43.000
6	47.25	29.50	61.000

Table 17: k-means based clustering of study program-institution ratio in private institution

Cluster	Number of member	Average score	Provinces list
1	4	2.64	Jambi, Bangka Belitung, West Kalimantan, North Kalimantan
2	10	3.42	Aceh, South Sumatera, Lampung, Central Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, South Sulawesi, West Sulawesi, Papua
3	8	3.95	Banten, North Sumatera, West Sumatera, Riau, Central Sulawesi, South-East Sulawesi, Maluku, Bali
4	7	4.98	DKI Jakarta, West Java, Central Java, Riau Island, West Papua, West Nusatenggara, East Nusatenggara
5	5	5.85	DI Yogyakarta, East Java, Bengkulu, Gorontalo, North Maluku

maximum score in Sumatera Region is 56% of the maximum score in Java Region, the minimum score in Sumatera Region is only 2.9% of the minimum score in Java Region. It means that there is big inequality among provinces in Sumatera Region. Meanwhile, the inequality in East Region and little Sunda is very low.

The 8th analysis is the study program-institution ratio in public institution. This ratio is obtained by dividing the number of study programs in with number of institutions in the same province. The purpose is to analyze the public higher education institutions to running as study programs as many as possible. Table 15 shows the k-means based clustering result. Meanwhile, Table 16 shows the geographical similarity based clustering result.

Table 15 shows that the fifth cluster contains provinces from Java, Sumatera and Sulawesi. Provinces in Java Region are distributed into third cluster to fifth cluster. The first cluster contains provinces from Sumatera, Kalimantan and Sulawesi. All provinces in East Region are in the second cluster. Based on this result, provinces in Java Region tends to have medium to high study program-institution ratio. Provinces in East Region and little Sunda tend to have medium to low study program-institution ratio. Provinces in Sumatera, Kalimantan and Sulawesi are distributed from low to high ratio.

Fact in Table 16 strengthens the k-based clustering result as it is shown in Table 15. The additional analysis

is as follows. In public higher education study program-institution ratio aspect, wide inequality occurs in Sumatera and Sulawesi. The most equal region is East Region. In East Region, this ratio tends to low. So, government should reduce inequality in Sumatera and Sulawesi by supporting provinces with low study program-institution ratio to create many prospective study programs. Meanwhile, government also should support all public higher education institutions in East Region to open many prospective study programs.

The 9th analysis is the study program-institution ratio in private institution. The purpose is to analyze the private higher education institutions to running as study programs as many as possible. Table 17 shows the k-means based clustering result. Table 18 shows the geographical similarity based clustering result.

Table 17 shows that the study program-institution ratio in private higher education is generally low. All of the average scores are <10. This condition is very different to the public higher education institution where all of the average scores are higher than 15. Although, all of average scores are low, the disparity between the first cluster and the fifth cluster is narrow. It means that generally, private higher education institution faces similar problem: lack of resources to create many program studies.

This condition can also be seen in different views: most of private higher education institutions focus on popular study programs, so that, it is easy to obtain

Table 18: Geographical similarity based clustering of study program-institution ratio in private institution

Region	Average score	Minimum score	Maximum score
1	5.10	4.25	5.84
2	3.68	2.00	5.50
3	3.19	2.73	3.64
4	3.97	3.41	5.62
5	4.74	3.40	6.53
6	4.76	4.10	5.13

Table 19: k-means based clustering of number of students in public institution

Cluster	Number of member	Average score	Provinces list
1	6	9,083.50	Riau Island, Bangka Belitung, Central Kalimantan, North Kalimantan, West Sulawesi, West Papua
2	11	26,635.09	Banten, Jambi, South Sumatera, Bengkulu, Lampung, South Kalimantan, Gorontalo, Maluku, North Maluku, Papua, West Nusatenggara
3	12	58,907.50	Aceh, North Sumatera, West Sumatera, Riau, West Kalimantan, East Kalimantan, North Sulawesi, Central Sulawesi, South Sulawesi, South-East Sulawesi, Bali, East Nusatenggara
4	3	157,719.00	West Java, Central Java, DI Yogyakarta
5	2	342,067.00	DKI Jakarta, East Java

Table 20: Geographical similarity based clustering of number of students in public institution

Region	Average score	Minimum score	Maximum score
1	195.804	17.532	401.009
2	42.5420	4.9540	86.5460
3	28.4310	5.4110	45.2970
4	44.4270	5.3370	71.1300
5	19.9690	15.312	26.9730
6	40.9440	3.0436	49.1180

students and economic of scale can be achieved more easily. In the other side as shown in Table 15, public higher education does not have big problem in operating many study programs, although, these study programs are not so popular.

Based on the distribution of provinces among clusters, it is shown that provinces in Java Region are distributed in medium to high clusters. Provinces in Sumatera are distributed in all regions. Besides Java, the distribution is more equal.

The analysis in Table 17 is strengthened by the result in Table 18. The study program-institution ratio in all regions is low. But there are several notes. The ratio between Java and little Sunda is similar. Meanwhile, the ratio in Kalimantan is the lowest among other regions. It means that it will be better if government gives incentives to private higher education institutions in Kalimantan to open more study program.

The 10th analysis is the number of students in public higher education institution. Table 19 shows the k-means based clustering result. Meanwhile, Table 20 shows the geographical similarity based clustering result.

Table 19 shows that DKI Jakarta and East Java dominate the number students in public institution in Indonesia. The fifth cluster average score is more than twice the fourth cluster average score that consists of West Java, Central Java and DI Yogyakarta. Besides Banten, all provinces in Java dominate the student size in public higher education institutions in Indonesia.

Meanwhile, besides Java region, provinces in all regions in Indonesia are distributed in the first cluster to the third cluster. Unfortunately, the disparity between the average score in the first cluster is very wide compared with the average score in the fifth cluster. Based on this result, it is shown that equality in student size among provinces in Indonesia is very wide. This inequality also occurs between provinces in Java and provinces outside Java. The average score of the first cluster is only 2.6% of the average score of the fifth cluster.

The analysis in Table 19 is also strengthened by the result in Table 20. Table 20 shows that Java Region dominates the average score, minimum score and maximum score. Only little Sunda Region minimum score can exceed Java minimum score. Meanwhile in Java Region, disparity between minimum score and maximum score is very wide. The minimum score is only 4% of the maximum score. In Sumatera, the minimum score is 5.7%.

The 12th analysis is the number of students in private higher education institution. Different to the previous process, this process measures the capability of the private institution to obtain and to serve students. Table 21 shows the k-means based clustering result. Meanwhile, Table 22 shows the geographical similarity based clustering result.

Table 21 shows that the student body disparity in private higher education institution in Indonesia is wide too. The average score of the first cluster is 4.7%. Based

Table 21: k-means based clustering of number of students in private institution

Cluster	Number of member	Average score	Provinces list
1	16	30,102.56	Riau Island, Jambi, Bangka Belitung, Bengkulu, West Kalimantan, Central Kalimantan, North Kalimantan, North Sulawesi, Gorontalo, Central Sulawesi, West Sulawesi, South-East Sulawesi, Maluku, North Maluku, Papua, West Papua
2	10	78,895.40	Aceh, West Sumatera, Riau, South Sumatera, Lampung, South Kalimantan, East Kalimantan, Bali, West Nusatenggara, East Nusatenggara
3	2	223,553.50	Banten, DI Yogyakarta
4	3	359,450.00	Central Java, North Sumatera, South Sulawesi
5	3	638,930.33	DKI Jakarta, West Java, East Java

Table 22: Geographical similarity based clustering of number of students in private institution

Region	Average score	Minimum score	Maximum score
1	463.959	205.872	683.114
2	92.4260	6.66000	343.790
3	36.2910	4.27800	57.4760
4	81.6340	19.5070	314.704
5	31.8290	26.2630	45.8860
6	68.7520	62.5170	76.1810

on this condition, the inequality in student body in private institution is wider than the inequality in public institution. Meanwhile, although the disparity between the number of institution between the public one and the private one is wide, the disparity in student body is less wide.

Compared among cluster, provinces in Java are distributed from the third cluster to the fifth cluster. Meanwhile, North Sumatera and South Sulawesi are in the fourth cluster. Unfortunately, all provinces in East Region are in the first cluster. All provinces in the little Sunda are in the third cluster.

Result in Table 22 strengthens the result in Table 21. Java Region dominates student body in private institution in Indonesia. The good news is that in Java Region, the disparity between minimum score and the maximum score in private institution is not as wide as the disparity in public institution. In private institution, the student body minimum score is only 30% of the student body maximum score. In Sumatera Region, the disparity is very wide. In Sumatera Region, the minimum score is only 1.9% of the maximum score. Based on this fact in Sumatera, the disparity in private institution is wider than in public institution.

Geographical similarity based clustering result in public institution (Table 20) and in private institution (Table 22) make the analysis about private sector contribution is easier. Moeliodihardjo (2014) said that private institution contributes two third of student enrollment (Moeliodihardjo, 2014). Compared with the average score in these tables, the analysis is as follows. Based on these tables, it is shown that in Java, Sumatera and Kalimantan, the private institution contributes approximately two third of student enrollment, so that, this number is close to Moeliodihardjo's number (Moeliodihardjo, 2014). Meanwhile in Kalimantan, East

Region and little Sunda, private institution contributes far less than two third of student enrollment. This fact shows hypothesis that private sector is difficult to operate or generate revenue in these regions. Another hypothesis is the operational cost in these provinces is high.

The 13th analysis is the ratio between number of students and number of institutions in public institution. This ratio is obtained by dividing the number of students with the number of institutions in same province. The purpose of this analysis is to measure the efficiency and the health of the institution. By neglecting other parameters, bigger number of students in every institution means bigger capacity, health and continuity. But, too big student body can make the institution becomes overcrowded and it can make fiscal pressure too (Thomson, 2008; Johnstone *et al.*, 1998). Besides overcrowded, fiscal pressure can be seen in institutions which are lack of academic equipment and libraries and their faculties are underpaid (Thomson, 2008; Johnstone *et al.*, 1998). Table 23 shows the k-means based clustering result. Meanwhile, Table 24 shows the geographical similarity based clustering result.

Table 23 shows that DKI Jakarta is the only member in the fifth cluster. The disparity between DKI Jakarta and other cluster is wide. The average score of the fourth cluster is less than a half of the average score in the fifth cluster. Meanwhile, the first cluster average score is only 6.8% of the fifth cluster average score. Besides DKI Jakarta, provinces in Java are distributed in the third cluster and in the fourth cluster. Meanwhile, besides provinces in Java, the members of the third cluster or the fourth cluster come from Sumatera, Sulawesi, little Sunda and East Region. Unfortunately, provinces in Kalimantan are in the first cluster or in the second cluster. Based on this fact, public institutions in Java tend to have medium to high student-institution ratio.

Table 23: k-means based clustering of student-institution ratio in public institution

Cluster	Number of member	Average score	Provinces list
1	6	5,463.030	Riau Island, Bangka Belitung, North Kalimantan, West Sulawesi, Maluku, West Papua
2	11	11,713.29	West Java, Aceh, Lampung, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, South Sulawesi, Papua, Bali, East Nusatenggara
3	10	18,700.41	Banten, Central Java, East Java, West Sumatera, Riau, South Sumatera, Bengkulu, North Sulawesi, Gorontalo, North Maluku
4	6	31,736.26	DI Yogyakarta, North Sumatera, Jambi, Central Sulawesi, South-East Sulawesi, West Nusatenggara
5	1	80,201.80	DKI Jakarta

Table 24: Geographical similarity based clustering of student-institution ratio in public institution

Region	Average score	Minimum score	Maximum score
1	30,043	14,792	80,202
2	16,540	2,4770	29,264
3	10,093	5,4110	13,892
4	21,591	5,3370	43,089
5	10,010	7,1000	16,293
6	17,845	10,820	30,436

Table 25: K-means based clustering of student-institution ratio in private institution

Cluster	Number of member	Average score	Provinces list
1	5	686.920	Aceh, Bangka Belitung, Central Kalimantan, North Kalimantan, North Sulawesi
2	9	1,05,.65	West Sumatera, Jambi, South Sumatera, Lampung, West Kalimantan, East Kalimantan, West Sulawesi, Maluku, Papua
3	8	1,249.29	North Sumatera, Riau, Riau Island, South Kalimantan, South-East Sulawesi, Bali, West Nusatenggara, East Nusatenggara
4	6	1,519.44	Central Java, Bengkulu, Central Sulawesi, South Sulawesi, North Maluku, West Papua
5	6	1,899.16	DKI Jakarta, West Java, Banten, DI Yogyakarta, East Java, Gorontalo

Table 26: Geographical similarity based clustering of student-institution ratio in private institution

Region	Average score	Minimum score	Maximum score
1	1,875	1,640	2,234
2	1,051	416	1,536
3	1,016	795	1,249
4	1,296	642	1,788
5	1,276	1,035	1,545
6	1,234	1,202	1,251

Table 24 shows that the student-institution disparity between minimum score and maximum score in a region is various. Disparity in Sumatera is very high. In Sumatera, the minimum score is only 8.4% of the maximum score. Meanwhile, the lowest disparity is in East Region because the minimum score is only 43.57%. It means that equality among provinces in student-institution ratio in public institution is the highest. In average score the East Region and Kalimantan is region that the student-institution ratio is very low among other provinces.

The 14th analysis is the ratio between number of students and number of institutions in private institution. Table 25 shows the k-means based clustering result. Meanwhile, Table 26 shows the geographical similarity based clustering result.

Table 25 shows that the student-institution ratio in private institution is far below the same ratio in public

institution. The highest average score in private institution is still lower than the lowest average score in public institution as it is shown in Table 23. The lowest average score in private institution is 12.5% of the lowest average score in public institution. Meanwhile, the highest average score in private institution is only 2.4% of the highest average score in public institution. Fortunately, the equality in this ratio in private institution is better than in public institution. In private institution, the lowest average ratio is 36.1% of the highest ratio.

The better equality in private institution is also shown in the number of members in every cluster. There is not any cluster which its member is <5 provinces or more than ten provinces. This condition is different to the public institution. In Table 23, it is shown that the fifth cluster contains only one province and in other side, the second cluster contains 11 provinces.

Table 27: k-means based clustering of student-study program ratio in public institution

Cluster	Number of member	Average score	Provinces list
1	1	198.16	Bangka Belitung
2	1	211.81	West Java
3	11	255.20	South Sumatera, Bengkulu, Lampung, Central Kalimantan, South Kalimantan, Gorontalo, South Sulawesi, Maluku, Papua, West Papua, Bali
4	17	327.66	Banten, Central Java, DI Yogyakarta, East Java, Aceh, North Sumatera, West Sumatera, Riau, Riau Island, Jambi, West Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, West Sulawesi, North Maluku, East Nusatenggara
5	4	638.74	DKI Jakarta, Central Sulawesi, South-East Sulawesi, West Nusatenggara

Table 28: Geographical similarity based clustering of student-study program ratio in public institution

Region	Average score	Minimum score	Maximum score
1	411.64	211.81	985.28
2	293.48	198.16	384.46
3	285.59	267.07	316.70
4	375.58	233.38	582.28
5	287.71	236.66	378.91
6	368.44	239.60	498.95

Unfortunately, disparity between Java and non Java occurs. All provinces in Java except Central Java are in the fifth cluster. Only one outside Java province is in the fifth cluster. Fortunately, the equality among provinces in all outside Java Region is good. There is not any disparity between West Indonesia and East Indonesia.

Result in Table 26 strengthens the fact that Java Region is superior compared with other regions in student-institution aspect in private institution. This region is the highest in average score, minimum score and maximum score. Except Sulawesi, the Java minimum score beats all other regions maximum score.

Based on this fact, government should give incentives, stimulus or any other aids, so that, the private institution can increase its student body. Different to public institution where most of its revenue comes from government budget, private institution budget is depended on the tuition fee, so that, bigger student body means bigger revenue. In this case, government should give more attention in Sumatera, Kalimantan, Sulawesi because the minimum score in these regions is below 1,000 students per institution.

Student-institution ratio comparison between public institution and private institution is linear with Moeliodihardjo's research (Moeliodihardjo, 2014). Although, private sector contributes significantly in making higher education more accessible, the health and continuity of many private institutions is still questioned. Moeliodihardjo (2014) said that private institutions still face critical problem that the proliferation of institution produces many small institutions with small number of students. Fact in this research shows that in all regions and in all clusters, the student-institution ratio in private institution is far below in public institution.

The 15th analysis is the ratio between number of students and number of study programs in public institution. The purpose of this process is to analyze the average load of the study program in public institution. Table 27 shows the k-means based clustering result. Meanwhile, Table 28 shows the geographical similarity based clustering result.

Table 27 shows that there is not any correlation between region and the student-study program ratio. Provinces in Java Region are distributed into the second, the fourth and the fifth cluster. There is only one province in Java Region that is in the fifth cluster. Meanwhile in the fifth cluster, there are provinces from Java, Sulawesi and little Sunda. Most of provinces are concentrated in the third and the fourth cluster. There is only one province that is in the first cluster. Meanwhile, the disparity between the first cluster and the fifth cluster is not wide. The average score in the first cluster is 31% of the average score in the fifth cluster.

Result in Table 28 strengthens the analysis that student-study program ratio among regions is almost equal. The lowest average score is only 69% of the highest average score. Meanwhile, the lowest minimum score is 74% of the highest minimum score. The wide disparity only occurs in maximum score where the lowest maximum score is 32%. Based on this result in public institution, there is not any stigma about study program load between Java and non Java.

The 16th analysis is the ratio between number of students and number of study programs in private institution. Table 29 shows the k-means based clustering result. Meanwhile, Table 30 shows the geographical similarity based clustering result.

Table 29 shows that the equality in student-study program ratio in private institution is better than the same ratio in public institution. The lowest average score is

Table 29: k-means based clustering of student-study program ratio in private institution

Cluster	Number of member	Average score	Provinces list
1	3	204.09	Bangka Belitung, Central Kalimantan, North Sulawesi
2	6	244.30	Aceh, West Sumatera, Riau Island, North Maluku, West Nusatenggara, East Nusatenggara
3	8	286.85	East Java, South Sumatera, Bengkulu, Lampung, North Kalimantan, Maluku, West Papua, Bali
4	9	329.45	Central Java, North Sumatera, Riau, Jambi, East Kalimantan, Gorontalo, West Sulawesi, South-East Sulawesi, Papua
5	8	396.69	DKI Jakarta, West Java, Banten, DI Yogyakarta, West Kalimantan, South Kalimantan, Central Sulawesi, South Sulawesi

Table 30: Geographical similarity based clustering of student-study program ratio in private institution

Region	Average score	Minimum score	Maximum score
1	370.15	293.08	421.94
2	283.36	208.13	352.12
3	322.61	218.74	409.11
4	325.66	185.39	410.84
5	275.91	236.60	320.88
6	262.11	237.71	304.72

Table 31: k-means based clustering of number of lecturers in public institution

Cluster	Number of member	Average score	Provinces list
1	4	246.750	Riau Island, Bangka Belitung, North Kalimantan, West Sulawesi
2	9	761.220	Banten, Jambi, Bengkulu, Central Kalimantan, Gorontalo, North Maluku, Papua, West Papua, West Nusatenggara
3	10	1,462.40	Riau, South Sumatera, Lampung, West Kalimantan, South Kalimantan, East Kalimantan, Central Sulawesi, South-East Sulawesi, Maluku, East Nusatenggara
4	8	3,418.50	DKI Jakarta, DI Yogyakarta, Aceh, North Sumatera, West Sumatera, North Sulawesi, South Sulawesi, Bali
5	3	7,932.33	West Java, Central Java, East Java

51% of the highest average score. The private institution first cluster average score is higher than the public institution first cluster average score. Meanwhile, the private institution first cluster average score is lower than the public institution fifth cluster average score.

Table 29 also shows that the disparity between provinces in Java Region and outside Java Region is narrow. Provinces in Java Region are distributed into the third cluster to the fifth cluster. Meanwhile, there are provinces in Kalimantan and Sulawesi that are in the fifth cluster. In other side, there are provinces in Sumatera, Kalimantan and Sulawesi that are in the first cluster.

Result in Table 30 strengthens the analysis based on Table 29. The disparity among regions is narrow. The lowest average score is 70% of the highest average score. Java Region tends to have higher student-study program ratio rather than other regions. All other region maximum score is higher than Java minimum score.

The 17th analysis is the number of lecturers in public institution. Table 31 shows the k-means based clustering result. Meanwhile, Table 32 shows the geographical similarity based clustering result.

Table 31 shows that the fifth cluster contains three biggest provinces in Java. Provinces in East Region are in the second and the third cluster. Meanwhile, the disparity between the first cluster and the fifth cluster is very wide.

The average score in the first cluster in only 3.1% of the average score in the fifth cluster. The fourth cluster average score is less than half of the fifth cluster average score. The third cluster average score is less than half of the fourth cluster average score. The second cluster average score is approximately half of the third cluster average score. The first cluster average score is less than half of the second cluster average score. Based on this fact, government should reduce this public institution lecturer disparity.

Result in Table 32 strengthens the analysis of Table 31. The Java average score is significant higher than other regions average score. Nevertheless, the little Sunda minimum score is much higher than the Java minimum score. The inner region equality in East Region and little Sunda is the highest with the minimum score is 37% of the maximum score. Meanwhile, inner region inequality in Java, Sumatera and Sulawesi is very high with the minimum score is <10% of the maximum score.

The 18th analysis is the number of lecturers in private institution. Table 33 shows the k-means based clustering result. Meanwhile, Table 34 shows the geographical similarity based clustering result.

Table 33 shows that all provinces in the fifth cluster are in Java Region. Besides in the fifth cluster, other provinces in Java Region are in the second cluster. Similar

Table 32: Geographical similarity based clustering of number of lecturers in public institution

Region	Average score	Minimum score	Maximum score
1	5,557	641	9,725
2	1,634	231	3,317
3	1,067	221	1,604
4	1,668	232	3,232
5	830	533	1,418
6	1,755	1,074	2,859

Table 33: K-means based clustering of number of lecturers in private institution

Cluster	Number of member	Average score	Provinces list
1	9	840.3300	Bangka Belitung, Bengkulu, Central Kalimantan, North Kalimantan, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua
2	10	1,890.000	Riau Island, Jambi, West Kalimantan, South Kalimantan, East Kalimantan, North Sulawesi, Central Sulawesi, South-East Sulawesi, Papua, East Nusatenggara
3	7	4,004.710	Aceh, West Sumatera, Riau, South Sumatera, Lampung, Bali, West Nusatenggara
4	5	10,173.00	Banten, Central Java, DI Yogyakarta, North Sumatera, South Sulawesi
5	3	22,767.00	DKI Jakarta, West Java, East Java

Table 34. Geographical similarity based clustering of number of lecturers in private institution

Region	Average score	Minimum score	Maximum score
1	16,068	7,339	23,671
2	3,8320	371	12,600
3	1,4300	167	2,3250
4	2,8840	651	10,157
5	1,1840	922	1,5990
6	3,2470	2,717	3,6280

to the fact in public institution, disparity in private institution is also very wide. The first cluster average score is only 3.6% of the fifth cluster average score, so that, disparity in private institution is similar to the disparity in public institution.

Result in Table 34 strengthens the analysis from Table 33. Java Region dominates in number of lecturers. Meanwhile, the East Region is region with the lowest number of lecturers. Based on Table 34, the lowest average score is only 7.3% of the highest average score. The fact that the minimum score in Java is higher than 5,000 lecturers is very contrast with condition that the minimum score of most regions is <1,000 lecturers. Although, there are three regions with maximum score is above 10,000 lecturers, the maximum score of other three regions is below 5,000 lecturers.

By neglecting the method, it is shown that in number of lecturers, private institution outnumbers public institution. Unfortunately in fulltime lecturer who has doctoral degree, the number of lecturers in public institution is twice the number of lecturers in private institution (Anon, 2017). This fact shows that many students in private institution are taught by medium or low qualified lecturers. To solve this problem, government should give more doctoral degree scholarship or incentives for private institution lecturers, so that, quality disparity between private institution and public institution can be reduced.

The 19th analysis is the ratio between number of students and the number of lecturers in public institution. This ratio means the average number of students that are served by single lecturer. This ratio is very useful to measure the lecturer's load. This ratio must be moderate. Lower value means the lecturer's idle time is high. In the other side, higher value means the lecturer's load to serve students is too high, so that, it can reduce lecturer's time and resource to run research and developing teaching materials. This condition is called overcrowded and it is typical problem in higher education institution in developing countries (Thomson, 2008).

Higher student-lecturer ratio is also means larger class or large number of students in a class. This ratio is an important metric because number of students in a class affects the quality of learning environment (Ehrenberg *et al*, 2001; Cuseo, 2007; Hornsby and Osman, 2014). Although, class size is not the only one factor that affects student performance, large class is believed to correlate with low performance student (Hornsby, 2013). Table 35 shows the k-means based clustering result. Meanwhile, Table 36 shows the geographical similarity based clustering result.

Table 35 shows that the fifth cluster contains three provinces from different regions. Although, Papua are always in the low cluster in this ratio, it is in the highest cluster. It means that lecturer's load in public institution in Papua is very high and it is in the same cluster with DKI

Table 35: k-means based clustering of student-lecturer ratio in public institution

Cluster	Number of member	Average score	Provinces list
1	3	17.03	Central Kalimantan, Maluku, Bali
2	7	22.48	West Java, Aceh, South Sumatera, Bangka Belitung, South Kalimantan, South Sulawesi, West Sulawesi
3	6	25.15	Central Java, West Sumatera, Bengkulu, Lampung, North Kalimantan, North Sulawesi
4	15	29.89	Banten, DI Yogyakarta, East Java, North Sumatera, Riau, Riau Island, Jambi, West Kalimantan, East Kalimantan, Gorontalo, Central Sulawesi, North Maluku, West Papua, West Nusatenggara, East Nusatenggara
5	3	53.54	DKI Jakarta, South-East Sulawesi, Papua

Table 36: Geographical similarity based clustering of student-lecturer ratio in public institution

Region	Average score	Minimum score	Maximum score
1	37.50	22.83	90.15
2	26.53	21.45	32.09
3	25.52	18.90	32.61
4	27.72	21.59	36.58
5	26.54	15.02	33.89
6	26.01	17.18	32.52

Table 37: k-means based clustering of student-lecturer ratio in private institution

Cluster	Number of member	Average score	Provinces list
1	5	18.13	Aceh, Bangka Belitung, Central Kalimantan, North Sulawesi, West Nusatenggara
2	14	23.50	West Sumatera, Riau Island, Jambi, South Sumatera, Bengkulu, Lampung, South Kalimantan, East Kalimantan, North Kalimantan, Gorontalo, South-East Sulawesi, North Maluku, Bali, East Nusatenggara
3	8	27.77	DKI Jakarta, West Java, Banten, East Java, North Sumatera, Central Sulawesi, Maluku, Papua
4	5	30.35	DI Yogyakarta, Riau, South Sulawesi, West Sulawesi, West Papua
5	2	32.51	Central Java, West Kalimantan

Table 38: Geographical similarity based clustering of student-lecturer ratio in private institution

Region	Average score	Minimum score	Maximum score
1	29.24	26.52	32.27
2	22.99	16.50	30.18
3	25.08	18.08	32.75
4	26.03	19.50	30.98
5	26.91	21.94	29.53
6	21.35	18.62	23.01

Jakarta. Provinces in Java Region are distributed from the second cluster to the fifth cluster. Provinces in Kalimantan are distributed into the first cluster to fourth cluster. Provinces in Sumatera are distributed into the second cluster to the fourth cluster. Provinces in Sulawesi are distributed into the second cluster to the fifth cluster. Based on this fact, it can be said that in lecturer's load aspect in public institution, there is not any difference among regions in Indonesia. In large class perspective, large class phenomenon is potential to occur in three provinces in the fifth cluster.

Table 36 shows a little bit difference. In public institution, lecturer's load in Java tends a little bit higher than lecturer's load in other regions. This condition can be seen in average score part where the average score in Java Region is higher than 35. Meanwhile, lecturer's load in other region tends to be similar. The average score are higher than 25 and lower than 30. Except in Java Region, the minimum score and maximum score in other regions are similar too.

The 20th analysis is the ratio between number of students and the number of lecturers in private institution. Table 37 shows the k-means based clustering result. Meanwhile, Table 38 shows the geographical similarity based clustering result.

Table 37 shows that there is little bit difference in lecturer's load in private institution and in public institution. In private institution, the disparity between the lowest cluster and the highest cluster is narrower rather than that is in public institution. In private institution, the first cluster average score is 56% of the fifth cluster average score. Meanwhile in public institution as it is shown in Table 35, the first cluster average score is 32% of the fifth cluster average score.

In private institution, there is disparity between provinces in Java Region and provinces in other regions. Provinces in Java Region are distributed into the third cluster to the fifth cluster. Meanwhile, provinces in other regions are distributed more equally. Provinces in Sumatera and Sulawesi are distributed into the first cluster

to the fifth cluster. Provinces in Kalimantan are distributed in the first cluster, the second cluster and the fifth cluster.

Result in Table 38 strengthens analysis that lecturer's load in private institution is more equal rather than in public institution. Although, Java Region has higher average score and minimum score rather than other regions, the disparity is not wide. The average score of all regions is higher than 20 and lower than 30. Meanwhile, the Java maximum score is lower than Kalimantan maximum score.

Findings: This research has met the research purpose in giving better perspective and analysis of the higher education condition, quality and equality in Indonesia. This research also enriches and produces new understanding rather than just statistical data in the previous research (MRTHERI, 2017). In the previous research (MRTHERI, 2017), data order is presented by the provinces position from the West of Indonesia to the East of Indonesia. By using k-means method, data has been clustered, so that, the analysis will be easier. Better analysis has been done especially in finding disparity among cluster and similarity inside the cluster. By using geographical similarity based method, performance and disparity among regions can be analyzed better too. This research has also analyzed disparity between private institution and public institution.

Based on the analytical work, there are several findings about higher education condition: basically, private institution contributes higher number of students rather than public institution. Approximately, two third higher education students in Indonesia go to private institution, especially in high populated regions. This number is similar to the number from previous research (Moeliodihardjo, 2014). Meanwhile in low populated regions, private sector contributes less than that number. In both private institution and public institution, except North Sumatera and South Sulawesi, provinces in Java Region tends to dominates other regions in number of institutions, number of students, number of lecturers and number of study programs. Generally, private institution has less students rather than public institution and this disparity is very wide. This condition occurs in all clusters and in all regions.

Generally, private institution has lower study program per institution ratio rather than public institution. This disparity is very wide where in most of regions, the ratio in public institution is more than ten times of in private institution. Student-lecturers ratio in public institution tends to be higher than in private institution but this disparity is narrow.

CONCLUSION

The idea to clustering the government data by using k-means method has improved and enriched perspective and findings about higher education condition in Indonesia. In some aspects, disparity among regions is very wide. In other aspects, there is not any difference among regions, especially, Java Region with other regions. In this condition, analysis from k-means based clustering result can be made easier to find similarities between provinces in the same cluster. k-means based clustering process has helped to find how far the disparity or inequality that occurs in higher education in Indonesia. Meanwhile, geographical similarity based clustering process has provided result about disparity among regions. When there is significant disparity between Java Region and other regions, geographical similarity based clustering strengthens result that is provided by k-means based method. In some aspects, there is disparity between private institution and public institution. In other aspects, there is not any difference between private and public institution.

REFERENCES

Anonymus, 2014. Higher education in South-East asia: A policy digest. Elementa Leadership Limite, Collett Place, Latton, Swindon. http://webcache.googleusercontent.com/search?q=cache:iEP_WjfTmjYJ:www.elementaleadership.co.uk/knowledgeportal/pdf/Higher%2520Education%2520in%2520South-East%2520Asia%2520-%2520A%2520Policy%2520D

Aritenang, A., 2008. A study on Indonesia regions disparity: Post decentralization. SSRN. Electron. J., 1: 1-10.

Bhogan, S., K. Sawant, P. Naik, R. Shaikh and O. Diukar *et al.*, 2017. Predicting student performance based on clustering and classification. IOSR. J. Comput. Eng., 19: 49-52.

Cuseo, J., 2007. The empirical case against large class size: Adverse effects on the teaching, learning and retention of first-year students. J. Faculty Dev., 21: 5-21.

Durairaj, M. and C. Vijitha, 2014. Educational data mining for prediction of student performance using clustering algorithms. Intl. J. Comput. Sci. Inf. Technol., 5: 5987-5991.

Ehrenberg, R.G., D.J. Brewer, A. Gamoran and J.D. Willms, 2001. Class size and student achievement. Psychol. Sci. Publ. Interest, 2: 1-30.

Hornsby, D., 2013. Teaching large classes in higher education: Challenges and opportunities in a developing context. Proceedings of the Canadian International Conference on Education, June 25, 2013, University of Toronto, Canada, pp: 1-4.

Hornsby, D.J. and R. Osman, 2014. Massification in higher education: Large classes and student learning. Higher Educ., 67: 711-719.

Johnstone, D.B., A. Arora and W. Experton, 1998. The financing and management of higher education: A status report on worldwide reforms. Report No.19129, World Bank, Washington, DC.

MRTHERI., 2017. Higher education statistical year book 2017. Ministry of Research, Technology and Higher Education of Republic of Indonesia, Central Jakarta, Indonesia.

Moeliodihardjo, B.Y., 2010. Equity and access in higher education: The case of Indonesia. Master Thesis, University of Indonesia, Depok, Indonesia.

Moeliodihardjo, B.Y., 2014. Higher education sector in Indonesia. Proceedings of the International Seminar on Massification of Higher Education in Large Academic Systems, November 10, 2014, National University of Educational Planning and Administration, New Dlehi, India, pp: 1-14.

Ogwoka, T.M., W. Cheruiyot and G. Okeyo, 2015. A model for predicting student's academic performance using a hybrid K-means and decision tree algorithms. Intl. J. Comput. Appl. Technol. Res., 4: 693-697.

Storen, I. and A.M. Tremonte, 2013. Book review: Higher education and national development: Universities and societies in transition. Excellence Higher Educ., 4: 56-59.

THE., 2018. Asia University rankings. Times Higher Education, London, UK.

Thomson, A., 2008. Exploring the relationship between higher education and development: A review and report. Master Thesis, Guerrand-Hermes Foundation for Peace, Brighton, England.