

## **CLUSTERING BASED EVALUATION ON ELEMENTARY SCHOOL EDUCATIONAL SERVICE EQUALITY**

**Feri Wijayanto**

*Department of Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia  
Jalan Kaliurang Km.14,5 Sleman, Yogyakarta 55184  
Email : feri.wijayanto@uii.ac.id*

### **ABSTRACT**

*Dengan lebih dari 200 juta penduduk, Indonesia memiliki potensi besar dalam sumber daya manusia, namun pendidikan masih menjadi polemik di negeri ini. Ketimpangan pelayanan pendidikan adalah salah satu dari banyak masalah yang menjadi polemik antar wilayah. Dalam banyak kesempatan, banyak yang mengatakan bahwa Pulau Jawa memiliki infrastruktur yang jauh lebih baik daripada pulau-pulau yang lain. Hal tersebut terkait dengan fakta bahwa ibukota Indonesia berada di Pulau Jawa. Penelitian ini dilakukan untuk mengetahui adakah keterkaitan antara kualitas pelayanan pendidikan dan jarak dari ibukota negara ke daerah. Penelitian dilakukan dengan menggunakan metode clustering untuk melihat kesamaan kualitas pendidikan di 33 provinsi di Indonesia. Dari clustering yang dilakukan, dapat disimpulkan bahwa jarak ke ibukota hampir tidak berpengaruh pada kualitas pelayanan pendidikan di daerah tertentu dan diyakini bahwa kualitas pendidikan adalah fungsi yang tergantung pada variabel yang lebih kompleks.*

*Kata kunci : Kualitas Pelayanan Pendidikan, Clustering, Educational Data Mining, Pendidikan Indonesia.*

### **1. INTRODUCTION**

As a vast country with more than 200 million populations, Indonesia has a great potential in human resources. With a great number of human resources, ASEAN Economic Community (AEC) should be a golden opportunity in terms of work vacancies. However, there is a concern to deal with the agenda. The concern due to lack of confidence in the quality of human resources which is rooted in the lack of trust in education quality.

Speaking about the quality of education, it is supposed to be linear with the quality of educational services, which means the better quality of educational service brings better quality of educational output. And related to educational problem in Indonesia, it is widely discussed that regions in Java Island have better educational service quality than other regions in other islands but there are no research publications found to support the discussion.

Researches on comparison of educational quality among regions are important to give

a cornerstone to a discussion related to the Indonesia's Educational service quality. Moreover, they will be useful as basis for the next stages of the researches.

Without those kinds of publications could we confidently said whether a region farther from the capital (The City of Jakarta) has a worse educational service quality compared to a closer one? Although it is common to discuss about quality differences of education (between regions which are around capital and far from capital) there is a gap between the common knowledge we get from public discussion and the knowledge based on scientific evidence.

Based on the gap that have been mentioned, this research uses clustering method to group all of the 33 provinces in Indonesia into several clusters. Clusters will be created based on the similarity of educational quality indicators among provinces which had been assigned by the Ministry of Education and Culture. Based on the clusters which are formed, an analysis will be carried out in order to see the relation between regions distance to the seat of

Indonesian government in Jakarta and the quality of educational service in each region. So that in the end we can ensure whether farther region from the capital has worse educational services quality compared to the closer region or not.

## 2. LITERATURE REVIEW

Equity issue of educational quality and services are always become a spotlight in studies of Indonesian human resources development plan. That is why the quality of education is always be the main course of discussion due to the hope of creating high quality of human resources which will enhance the competitiveness of Indonesian people.

There are many articles, writings or publications which are discussing about education quality in Indonesia especially its inequality among regions either liputan6.com which take a picture of education in a remote area (SCTV, 2014) or Harjatanaya who inform four main issues in Indonesia's education (Harjatanaya, 2012).

In addition to the writings in media, there are several research publications which investigate the quality of education in Indonesia. One of them was held in the province of Papua (Indubri et al., 2013). Darmawan also investigated the management of infrastructures in order to escalate the quality of education (Darmawan, 2014). Several years before, Salirawati observed the implementation of competence based curriculum including the problems which accompanied it (Salirawati and Sari, 2002). Those three researches had tried to see the quality of education in Indonesia and the instruments to achieve the quality.

On a different angle, Fadloli used geographic information system (GIS) to categorize the quality of education in order to manage the state elementary school in Widodaren district, Ngawi (Fadloli, 2014). Other than that, Slamet develop an information system for school categorizing

Table 1. .

as a base of student's choice in Semarang (Slamet, 2014). Both researches use information technology as a tool which can support education world.

Actually, the government as the key actors of educational quality improvement had published data and survey results related to educational report. Central Bureau of Statistics had published a survey result on elementary education infrastructure in 2011 (Statistik, 2011). Ministry of Education and Culture was also published a compendium of education data (Sidiknas, 2014). Unfortunately, the very last data which are published was on academic year of 2011-2012 and based on that limitation, those data which will be used as the research data.

Basically there are many researches and studies had been done related to education quality improvement, yet the research which are related to data presentation using data mining approach especially about the relationships between the distance and the quality of education services not many occurred yet. So that this paper positioned as a preliminary research about the educational services quality in Indonesia which is pointed to the relation between regions distances to the capital and the quality. As a preliminary research, the data in this research is limited to those data about elementary schools.

## 3. MATERIAL & METHODOLOGY

Data which are used in this research are taken from compendium which was published by The Ministry of Education and Culture (Sidiknas, 2014) and also from survey report of Central Bureau of Statistics on academic year 2011-2012 (Statistik, 2011). These data are the newest comprehensive recapitulation of educational data in Indonesia which was published by the Ministry of Education and Culture of Indonesia. Indicators which were used as variables in this research are shown in .

Table 1 consists of variables along with corresponding mission categories, units and their standard values (Kebudayaan, 2013).

Table 1 are the ideal values for each variable which are desired to be achieved in every region in Indonesia based on the Ministry of Education and Culture decree.

This research was begun by collecting data on all of those variables for 33 provinces in Indonesia. So there are 33 x 6 matrices of Indonesian educational data

The standard values which mentioned on .

which will be processed during the research. The geographical location of 33 regions which are numbered correspond with data source sequence are shown in Figure 1. After having collected the data, normalization will be conducted and followed by the use of clustering method.

Table 1. Research variables

MISSION	INDICATOR	UNIT	STANDARD
<b>K1</b>	Students/School Ratio	Students	166
	Students / Classroom Ratio	Students	28
	Classroom / room ratio	Classroom	1
<b>K3</b>	Qualified Teachers Percentage	Percentage	100
	Students / Teachers Ratio	Students	18
	Proper Classroom Percentage	Percentage	100



Figure 1. Physical Location of Each Data Object.

The problem on normalization phase is that in those standard only two which have percentage type, four other standards are ratio type. It is easy to deal with percentage type standard but on contrary it is harder to deal with ratio type. Although basically min-max normalization can be implemented on almost all of variable types, but differ with percentage type which already has a range

value (0-1), ratio type has a fixed value. It means that the value of ratio type will not in a range but rather vary either below fixed value ratio or above. So when this type of variable is normalized, it will lose its meaning of quality which the best quality is on the fixed value of ratio.

Min-max normalization as a simple transformation was chosen to transform the

data set (Han et al., 2006). It performs a linear transformation on the original data into range [0,1] by computing.

$$X_n = \frac{X_0 - X_{min}}{X_{max} - X_{min}} \dots\dots\dots(1)$$

where,

- $X_n$  = new value for variable X
- $X_0$  = current value for variable X
- $X_{min}$  = minimum value in data set
- $X_{max}$  = maximum value in data set

The next important phase is estimating the number of cluster. The phase basically will play a significant role to improve the performance of the clustering method. In this research, NbClust (a package of an R programming language) (Charrad et al., 2014) was used in order to determine the number of clusters. The package provides 30 indices for determining the number of clusters and proposes to use the best clustering scheme from the different results obtained by varying all combinations of number of clusters, distance measures, and clustering methods.

Due to number of objects which only 33 and the purpose of this research which is to seek evidence of relation between educational service qualities and the number of clusters will be best around 3 to 7. Either too small or too many number of clusters will cause difficulties to compare the distance among regions. Based on that reason the number of cluster will be the number which was proposed by NbClust among those five numbers (3, 4, 5, 6, and 7).

By applying euclidean as the distance measurement method and complete hierarchical clustering as the clustering method, NbClust propose 3, 4 or 7 as the number of clusters. Based on the proposed number of cluster, Indonesian territory division is shown on figure 2 with Jakarta as the center of the circle partition. Figure 2 (a), (b) and (c) shows three, four and seven linear division of territory respectively.

Hierarchical clustering had been chosen in order to know all relation among the data objects. With agglomerative hierarchical clustering method, a dedrogram will be obtained and explain the relation desired. The clustering process was implemented by using agnes function in cluster library of R programming (Struyf et al., 1996).

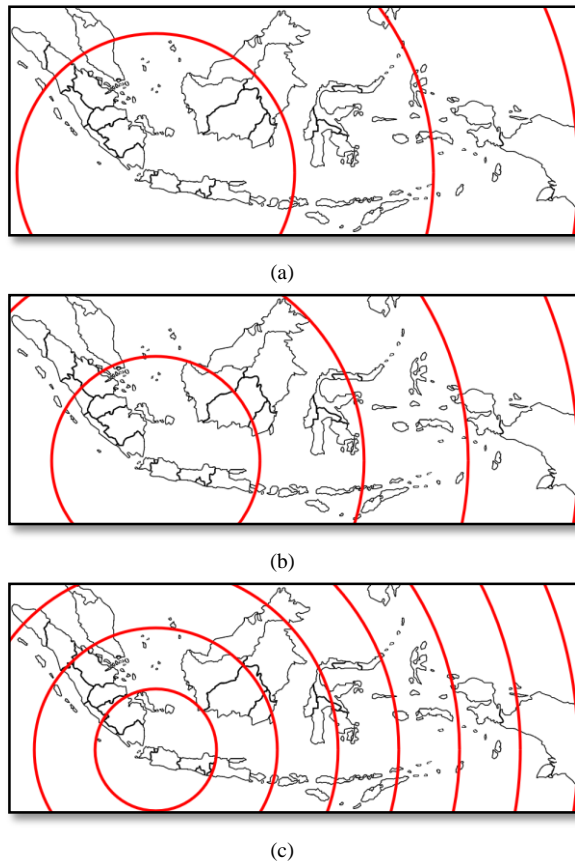


Figure 2. Linear Division of Indonesian Territory.

In order to see the relation between educational service quality levels with distance from the capital, the clustered regions then mapped on to a geographical map of Indonesia with different color. Afterwards the distribution of the regions with same level of educational quality will be seen on the map.

As long as the quality of regions has not been measured and declared which region has a better quality than the others, clustering cannot do much except establish a cluster. Clustering cannot analyze which

cluster is better and which is worse. In order to deal with the quality level, sum of squared error (SSE) of region's values of variables compared to standard values of variables.

$$QV(SSE) = \sum_{i=1}^6 (s_i - v_i)^2 \dots\dots\dots(2)$$

where,  
 QV = Quality value  
 s<sub>i</sub> = Standard variable value  
 v<sub>i</sub> = Region's variable value

Equation (2) is used to aim the deviation with the standard values which have been considered by the government. Simply mentioned that if standard have been issued the best quality is the closest to standard. This method put aside the different interpretation between values which are above and below standard ratio then treat them equally.

**4. RESULT AND DISCUSSION**

As a result of clustering process, the data are grouped into several clusters which indicates their similarities. Figure 3 shows the cluster dendrogram with three levels rectangles which separate the dendrogram into chosen number of clusters – 3,4 and 7 from top respectively.

Depth of the dendrogram describes the process of cluster establishment. The deepest (bottom) pair of dendrogram indicates the very first cluster which were established. It means that the most similar regions in terms of educational service quality are Jambi and Bangka Belitung (12 and 14 respectively). While the higher (upper) part of tree indicates the very last cluster, which means Riau Islands is the last region that joins the agglomerative tree. It also means that Riau Islands has the most dissimilarities among all regions in terms of educational service quality.

After having the clusters, their members were mapped to an Indonesian map in order to visualize the distribution of the cluster members. Figure 4 visualize the distribution of the cluster members through map on three

different numbers of clusters – 3, 4 and 7 respectively.

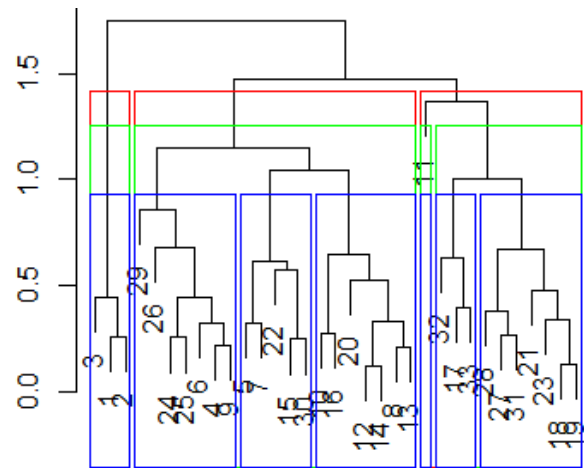


Figure 3. Cluster Dendrogram With Rectangles.

Figure 4 (a) indicates distance (radius) change compare to Figure 2 (a) on three number of clusters. While on Figure 2 (a) the distance is divided linearly with Jakarta as the center, Figure 4 (a) change the radius of the circle become smaller due to clustering result. As a center, Jakarta has only two other members in its cluster which reduce the radius a lot. Other than that, in the second distance area there are more than one clusters which are located their members there.

The first important fact on the Figure 4 (a) are those three regions in Kalimantan which are closer to capital (West Kalimantan, Middle Kalimantan and South Kalimantan) are more similar to Papua rather than with other regions in Java. The second is Gorontalo which are located quite far from the capital compare to other regions in Sulawesi instead of being similar to Papua, Gorontalo is more similar to other regions in Java.

As it happened on (a), Figure 4 (b) is also indicates distance (radius) change compare to Figure 2 (b) on four number of clusters. As a center and same as (a), Jakarta constantly has only two other members in its cluster. The new cluster is formed with only one member, which is Riau Islands. If observed in the dendrogram on Figure 3,

Riau Islands with number 11 actually has less similarity with other data.

There is not a big change from Figure 4 (a) to (b). A change which is clearly seen is an establishment of a new cluster which only consist of Riau Islands as its member.

The same phenomenon happens on Figure 4 (c) which is also indicates distance (radius) change compare to Figure 2 (c) on seven number of clusters. The phenomenon is rather complicated when the clusters members are mapped on the map and wanted to draw area for each cluster.

Figure 4 (c) which are represented seven number of clusters experience quite large change. Sumatera which is formerly become one cluster split up into two. Kalimantan, Java, and Sulawesi also added by one new cluster each. The first important fact shows that Special District of Aceh which is located on the farthest west side of Sumatra has a quite similar quality with Special District of Yogyakarta which is located on the center of Java. Second fact is that southern side regions of Sulawesi have more similar quality to those regions in Java compare to those in Kalimantan. And the third fact is that West Kalimantan is more similar to Papua in terms of educational service quality rather than regions in Java.

Basically those facts mean that the elementary school service quality are not merely linier function of distance but more complicated than that. The simplest prove that West Kalimantan which actually close to Jakarta has similar service quality with West Papua which are the farthest.

Aside from the region's quality mapping, by computing the QV (2) shocking fact was resulted. When using the standard variable values as the base, the regions with green color (Special district of Jakarta, West Java and Banten) are three regions with highest value of QV (SSE). It means that those three regions have the highest error to standard values among all of the regions. Shockingly, the highest error means the worst quality and that means regions with green color are the worst among all of region.

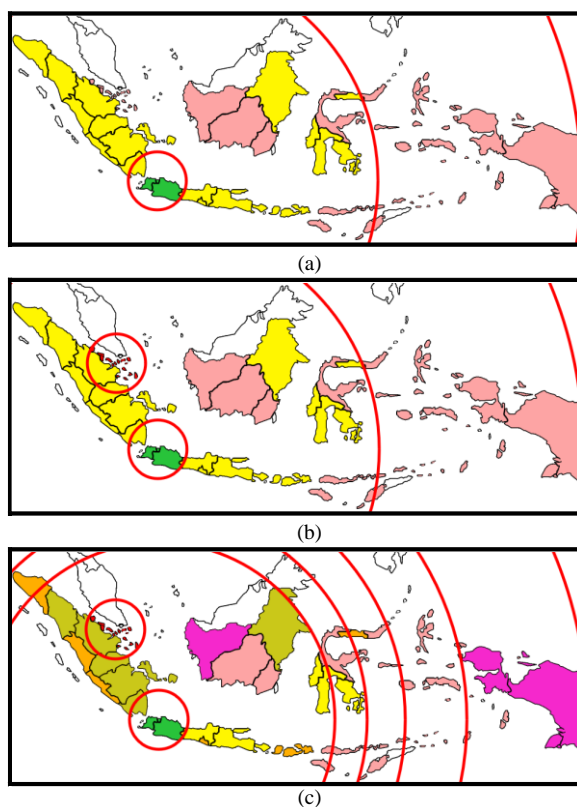


Figure 4. Distribution of The Cluster Members Through Geographical Map.

## 5. CONCLUSION

Based on the result, it is clear that distance from the capital is not a main parameter which give influence on inequality of educational service quality in elementary school level. It also means that the reason of a certain region having good or bad quality of educational services is not merely because its location from the capital. Most of regions in Kalimantan (Borneo) basically have more similar quality with regions in Papua than those in Java despite the location.

Shockingly, this research also found a contradictory result which mention Jakarta, West Java and Banten are the worst three in terms of educational services quality. In this research, the educational services quality is represented by the six research variables which are listed on Table 1.



Aside from the result. Although it is clear that distance is not influenced the quality, clustering cannot be used to conclude clearly that members from a certain cluster better than other cluster. Although there are not perfectly match between two results which have been obtained, it still can be accepted because those three regions (Jakarta, West Java and Banten) is in the same cluster.

So actually there are still lot of work to do related to unveiling the problem around inequality of educational service quality. Not only working with the data, but also new parameters which are related to the educational service. Further than that, developing the method in order to analyze and compare the quality among regions. Hopefully later on, the core problem will be unveiled and solved. And the next dream will be the equality of educational service among all regions in Indonesia.

The shortcoming of this research are the lack number of variables which was used. Further than that, deeper analysis of the variables is needed in order to interpret the values. The usage of min-max normalization is not effective on ratio type variables in this research context and need to use another method.

This research contributes an academic evidence of distance influence on educational service quality in order to answer what people have discussed and believed. It proves that distance from capital almost has no effect on inequality of educational service quality. Further than that, this research is positioned as a preliminary research on educational service quality by using data mining as a point of view. In addition, this research also gives a contribution on providing an evidence of min-max normalization weakness when it comes to ratio type attribute.

## REFERENCES

- Ayuliana, Teknik Pengujian Perangkat Lunak – Black Box Testing [online], 2012.
- Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A., NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, Vol. 61, No. 6, Pp: 1–36, 2014.
- Darmawan, B., Manajemen Sarana Dan Prasarana Dalam Meningkatkan Kualitas Pendidikan. *Jurnal Pelopor Pendidikan*, Vol. 6, No. 2, 2014.
- Fadloli, N., Pemetaan Kualitas Pendidikan dengan Menggunakan Sistem Informasi Geografis (SIG) untuk Pengelolaan Pendidikan Dasar Negeri Kecamatan Widodaren. *SKRIPSI Jurusan Geografi-Fakultas Ilmu Sosial*, 2014.
- Greenberg, S., Carpendale, S., Marquardt, N., and Buxton, B., The Narrative Storyboard: Telling a Story About Use and Context Over Time. *Interactions*, Vol. 19, No. 1, Pp: 64–69. 2012.
- Han, J., Kamber, M., and Pei, J., *Data Mining: Concepts and Techniques*. 2nd ed. Morgan Kaufmann Publishers Inc. San Francisco. 2006.
- Harjatanaya, T.Y., Pendidikan Indonesia, Sebuah Evaluasi [online]. *Edukasi*. 2012. Available: <http://nasional.kompas.com/read/2012/05/02/13011096/Pendidikan.Indonesia..Sebuah.Evaluasi> [Accessed 2014-2-20], 2012.
- Indubri, P.Y., Idrus, M.S., Salim, U., and . D., *Manajemen Pendidikan Sekolah dan Kualitas Pendidikan (Studi pada Kualitas Pendidikan di Provinsi Papua)*. *Jurnal Aplikasi Manajemen*. 2013.
- Kebudayaan, K.P.D., *Infografi Pendidikan Tahun 2011/2012*. Jakarta. 2013.

- Salirawati, D. and Sari, R., Peningkatan Kualitas Pendidikan Melalui Penerapan Kurikulum Berbasis Kompetensi, Antara Harapan Dan Kenyataan. Seminar Nasional dan Munas IKA UNY. Yogyakarta. 2002.
- SCTV, T.L. 6, Potret Pendidikan Daerah Terpencil [online]. Available: <http://news.liputan6.com/read/2044900/potret-pendidikan-daerah-terpencil>, 2004.
- Sidiknas, Buku Ringkasan Data Pendidikan [online]. Available: <http://kemdikbud.go.id/kemdikbud/node/2322>, 2014.
- Slamet, M., Sistem Informasi Peta Digital Mapping Sekolah di Kota Semarang dengan Metode Naive Bayes Classifier Sebagai Dasar Peminatan Siswa. Skripsi, Fakultas Ilmu Komputer. 2014.
- Statistik, B.P., Peta Tematik Infrastruktur Pendidikan Dasar 2011. Badan Pusat Statistik. Jakarta. 2011.
- Struyf, A., Hubert, M., and Rousseeuw, P.J., *Clustering in an Object-Oriented Environment*. Journal of Statistical Software, Vol. 1, No. 4. 1996.