Input-Support-Output Model Evaluation Using Clustering Analysis on Indonesia High School Dataset

Feri Wijayanto
Department of Informatics, Faculty of Industrial Technology,
Islamic University of Indonesia, Yogyakarta 55584, Indonesia.
feri.wijayanto@uii.ac.id

***Abstract*—Input-output model has been widely used in many research areas even in educational research. A previous research has proposed an adjusted input-support-output model to evaluate the quality of education development performance in Indonesia. Even though the previous research has found that the proposed model could explain 88% relation of input, support and output on each province when it was implemented on elementary school dataset, it is important to implement the model in other education level dataset to verify its performance. In this research, clustering analysis was used to cluster each group of the datasets of junior high school and senior high school prior to be mapped and simulated using the model. The results of the analysis of the model performance showed a decrease to 60.6% and 57.58% when it is implemented to junior high school and senior high school datasets respectively.**

***Index Terms*—Input-Output Model; Clustering; Education; Education Data Mining.**

I. Introduction

Input-output model has been widely used in many research areas. From economics [1], the model has been adapted to education [2] area, especially for evaluating the effectiveness of education process either in the scope of schools [3]–[5] or in the larger scale [6]. As a black box evaluation model, this model can be used to deliver the general overview of a problem.

Due to the condition of the Indonesian education system, which is changing fast and greatly in every government period, it is important to have many ways to obtain an overview from many angles of the current condition of education. Therefore, using similar cluster analysis [7], this research aimed to evaluate the input-output model by comparing the results of the clustering analysis at elementary school level. If the model could explain the relation of input - support - output well using different datasets, it could be considered as another way to observe the performance of education in Indonesia.

Based on an ongoing publication [7], it was found that the proposed input – support – output model could be used to explain the education performance in Indonesia's elementary schools among provinces by using a cluster analysis. From thirty-three provinces observed in Indonesia, the model could explain the input – support – output relation of 88% provinces. According to the cluster analysis, it implies that all of the four provinces have an unexpected behavior, which could not be explained clearly yet.

Witnessing the model that has been implemented successfully on elementary school data [7], it would be necessary to evaluate the performance quality of higher-level education using the same model. Based on the necessity to evaluate the model, this paper was designed to contribute to the evaluation of adopting this model using different datasets namely, from high school level (junior high school and senior high school) datasets.

II. Literature Review

Since education has already become an integral part in the development of a country, many researches have been conducted in the area of education to develop, investigate [8]–[11] or even enhance [12] the education system itself. Because of the importance of education, the former President of South Africa, Nelson Mandela, addressed it as a powerful weapon to change the world [13].

Indonesia already has its blueprint, which has been implemented since 2005 and it will be continued until 2025 to educate Indonesian people as intelligent and competitive persons. Table 1 shows several indicators used to evaluate the education performance periodically. The indicators have been categorized into three groups: support, output and input. The indicators have also been equipped with a national standard set by the Ministry of Education and Culture. The national standards are listed on JHS (Junior High School) and SHS (Senior High School) columns.

Table 1
Education quality indicators

Category	Indicators	Unit	JHS	SHS
Support	Students/School Ratio	Students	288	384
Support	Students/Class Ratio	Students	32	32
Support	Class/classroom ratio	Classroom	1	1
Support	Qualified Teachers	Percentages	100	100
Support	Students/Teachers Ratio	Students	18	19
Support	Proper Classroom	Percentage	100	100
Output	Completion Rate	Percentage	100	100
Output	Repetition Rate	Percentage	0	0
Input	Dropout Rate	Percentage	0	0
Input	Enrollment Rate	Percentage	100	100
Input	Gross Enrollment Rate	Percentage	100	100
Input	Transition Rate	Percentage	100	100

The educational data used in this research were based upon the statistical data, issued by The Ministry of Education and Culture and Central Bureau of Statistic in the academic year of 2011-2012 [14,15]. The earlier data were taken to tailor with the previous research [7] hoping to gain a consistent result.

A. *Input - Support - Output Model Overview*

Input-output model is a common model used and adapted to model a certain process. The model has been used in economics [1] and adapted to many areas, including in education [2]. In the area of education, the input-output model has been used for many purposes. Some researches used the model to model the education process [4,16] or even to evaluate the efficiency and effect of input and process on the education output [3,6,17].

In the previous research based on the commonly used input-output model, an adjusted input-support-output model was developed to evaluate the education performance quality among provinces in Indonesia [7]. Figure 1 shows the adjusted model, which include the indicators related to the model component. The model illustrates the relationship among the input, output, process, support and outcome. However, in this research, the scope is limited to input, output and the support with the assumption that the education process is linier or similar in all provinces.

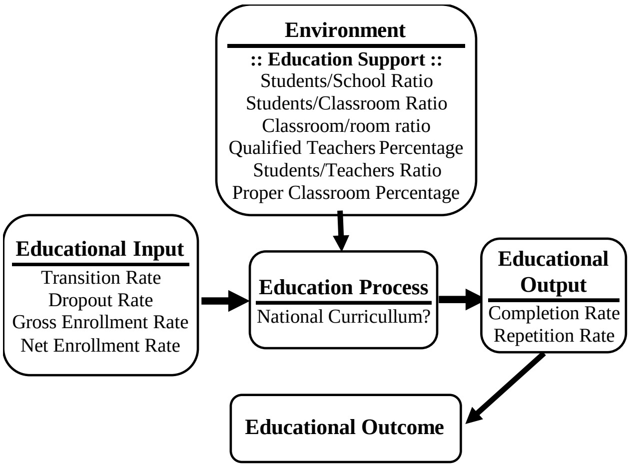

Figure 1: Adjusted input - support - output model [7]

Considering that not every province has colleges, the transition rate from senior high school to college cannot be presented based on province. It means that the transition rate data cannot be used for the analysis, and the indicators for senior high school input dataset will exclude the transition rate data.

III. Material and Methodology

Based on the research objective, the datasets implemented to the model are drawn from the high schools. The data have been taken from the compendium as released by Ministry of Education and Culture in the same year of the previous research [7]. The data were then served as the indicators as shown in Table 1.

Common clustering analysis methodology was used as shown in Figure 2. The research started by collecting the educational data. Transformation and normalization came after the datasets were collected. Then, the estimation of the number of clusters needed to be simulated was derived to gain the optimum clusters that should be created.

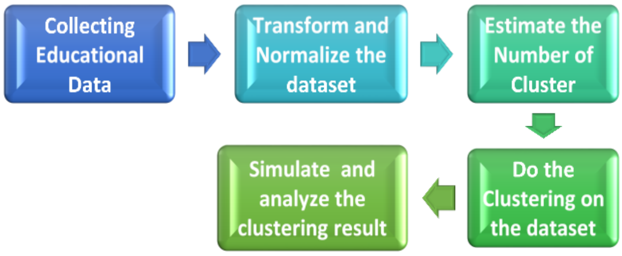

Figure 2: Research Methodology

In order to do the clustering, there are many algorithms which could be used. Based on the survey that had been made, k-Means is one of the most popular clustering algorithms [18] and have been used in many researches in many fields [19]–[21]. But based on the size of the datasets and the purpose of this research, which is to observe the hierarchical relation among objects [22], hierarchical clustering algorithm was used. After the clustering, the last part of the research was the analysis of the clustering result based on the proposed input – support – output model.

The high school datasets used consist of junior and senior high school datasets. Each dataset consists of several indicators which were then categorized into three groups of indicators: input, education process support and output. Therefore, at every level of high school, there would be three matrices of dataset, matrices of input, support and output. Input dataset that consisted of four indicators to be formed into 33x4 matric, educational support would be formed into 33x6 matric and output into 33x2 matric.

Data transformation and normalization have provided a significant impact on the cluster creation process. A similarity among objects in the hierarchical clustering was highly dependent upon the euclidean [23] distance used for computation. For several indicators with a ratio type - not percentage, simple min-max normalization was used to normalize the data [23]. Those indicators which needed to be normalized include student – school ratio, student – classroom ratio, classroom – room ratio and student – teacher ratio. Cost – benefit [24] transformation was implemented to level the indicators values. This transformation was required because some indicators were better when their value was higher in contrast to some indicators. In this research all indicators were made better if the values were smaller.

The estimated number of clusters was needed to analyze the number of groups that should be made based on the similarity of education quality among provinces. To compute the estimated number of clusters, NbClust of R library package [25] was used. This package applied 30 algorithms in finding the estimated number of clusters. The optimum number of clusters was determined by using the majority rule, meaning to choosing the number of clusters recommended by the most methods.

After the number of cluster was obtained, the hclust method of stats [26] R library package was implemented to find the members of each cluster. The last step of this research was the analysis on the clustering result of the junior and senior high school datasets and simulated the cluster's members using the proposed model. The quality level of each cluster was computed by using the average SSE (sum of squared error) within cluster. The error was obtained by computing the distance of the object to the national standard.

IV. RESULTS AND DISCUSSION

By using NbClust [25], the optimum number of cluster has been obtained for each dataset either junior high school or senior high school. For input, support and output datasets of junior high school, it has been obtained that the optimum number of clusters were 6, 3 and 3 respectively, as shown in Figure 3, 4 and 5. The red borders surround the members of each cluster.

By using the average SSE within cluster, the quality of each cluster on each dataset could be obtained and then sorted based on the quality levels. Table 2 shows the comparison map of junior high school input, support and output quality level of every province in Indonesia. The value on the input, support and output column indicates in which cluster where a province lies. The cluster number has been sorted by the quality level, the lower the cluster's number means the better quality it has.

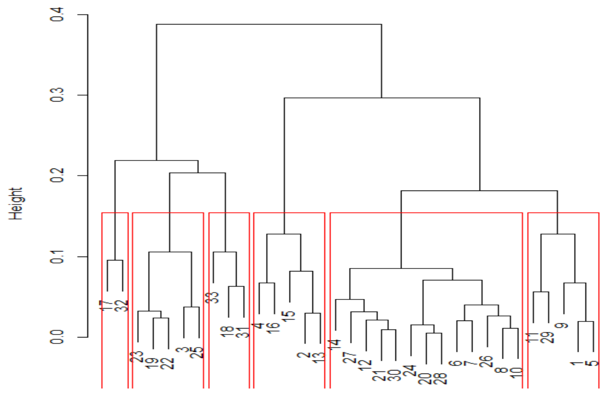

Figure 3: Junior high school input dataset dendogram

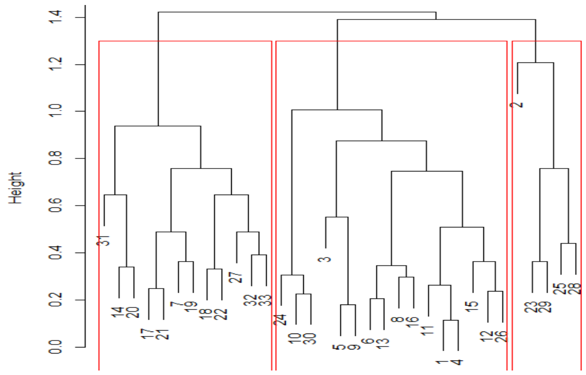

Figure 4: Junior high school support dataset dendogram

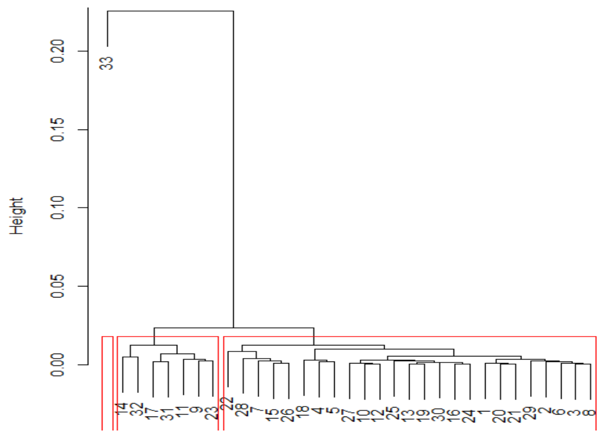

Figure 5: Junior high school output dataset dendogram

Table 2
Junior high school clusters comparison

No.	Provinces	input	support	output
1	Special District of Jakarta	1	1	1
2	West Java	3	3	1
3	Banten	4	1	1
4	Central Java	3	1	1
5	Special District of Yogyakarta	1	1	1
6	East Java	2	1	1
7	Special District of Aceh	2	2	1
8	North Sumatera	2	1	1
9	West Sumatera	1	1	2
10	Riau	2	1	1
11	Riau Islands	1	1	2
12	Jambi	2	1	1
13	South Sumatera	3	1	1
14	Bangka Belitung	2	2	2
15	Bengkulu	3	1	1
16	Lampung	3	1	1
17	West Kalimantan	6	2	2
18	Central Kalimantan	5	2	1
19	South Kalimantan	4	2	1
20	East Kalimantan	2	2	1
21	North Sulawesi	2	2	2
22	Gorontalo	4	2	1
23	Central Sulawesi	4	3	1
24	South Sulawesi	2	1	1
25	West Sulawesi	4	3	1
26	Southeast Sulawesi	2	1	1
27	Maluku	2	2	1
28	North Maluku	2	3	1
29	Bali	1	3	1
30	NTB	2	1	1
31	NTT	5	2	2
32	Papua	6	2	2
33	West Papua	5	2	3

As for senior high school datasets, NbClust [25] obtained 3, 5 and 6 as the optimum number of cluster for input, support and output dataset respectively. Those numbers are illustrated in Figure 6, 7 and 8 respectively.

Similar with junior high school datasets, by using the average SSE within cluster, the quality of each cluster on each dataset was obtained. Table 3 shows the comparison map of senior high school input, support and output quality level of every province in Indonesia. The values on the input, support and output column indicate in which cluster where a province lies. Similar with the junior high school dataset clusters, the cluster number has been sorted by the quality level in which the lower the cluster's number means the better quality.

Provinces with yellow shading in Table 3 show the phenomenon, which could not be explained clearly yet. For example, West Java with level 2 input and supported by level 3 support could result in a level 1 output. Another unexpected fact is Riau Islands with level 2 input and supported by level 1 support but only resulting in a level 2 output. Overall, the model 7 has only succeeded to explain 57.58% fact among provinces.

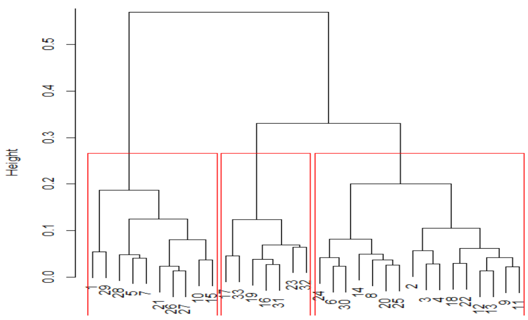

Figure 6: Senior high school input dataset dendogram

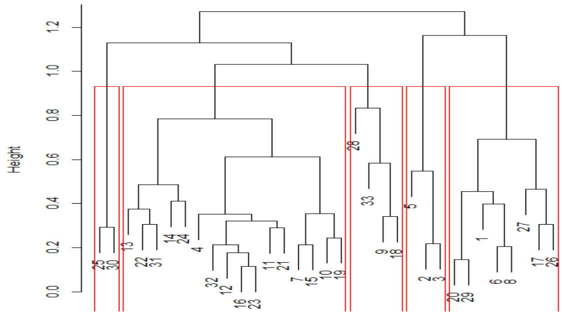

Figure 7: Senior high school support dataset dendogram

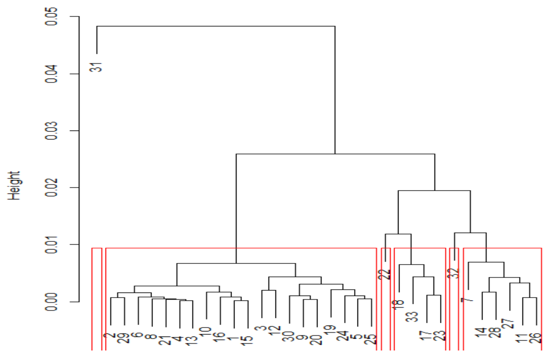

Figure 8: Senior high school output dataset dendogram

Table 3
Senior high school clusters comparison

No.	Provinces	input	support	output
1	Special District of Jakarta	1	2	1
2	West Java	2	3	1
3	Banten	2	3	1
4	Central Java	2	1	1
5	Special District of Yogyakarta	1	3	1
6	East Java	2	2	1
7	Special District of Aceh	1	1	2
8	North Sumatera	2	2	1
9	West Sumatera	2	4	1
10	Riau	1	1	1
11	Riau Islands	2	1	2
12	Jambi	2	1	1
13	South Sumatera	2	1	1
14	Bangka Belitung	2	1	2
15	Bengkulu	1	1	1
16	Lampung	3	1	1
17	West Kalimantan	3	2	3
18	Central Kalimantan	2	4	3
19	South Kalimantan	3	1	1
20	East Kalimantan	2	2	1
21	North Sulawesi	1	1	1
22	Gorontalo	2	1	5
23	Central Sulawesi	3	1	3
24	South Sulawesi	2	1	1
25	West Sulawesi	2	5	1
26	Southeast Sulawesi	1	2	2
27	Maluku	1	2	2
28	North Maluku	1	4	2
29	Bali	1	2	1
30	NTB	2	5	1
31	NTT	3	1	6
32	Papua	3	1	4
33	West Papua	3	4	3

V. Recommendation and Future Work

In summary, this research found some findings regarding the model or even the datasets, which were used in the analysis process. Based on the result of the cluster analysis, it is found that the output quality either junior high school or senior high school are quite similar. The similar pattern could be seen in Figure 5 and 8. Those figures show that even though the datasets have been grouped into more than one cluster, the distances among clusters are low.

The second finding is based on the input-support-output model's hypothesis, in which it is hypothetically assumed that input and support parameters have a direct influence to create a better output. Even though the model could explain better on elementary dataset (with 88% phenomenon succeed to be explained), it is found that only slightly above 50% phenomenon could be clearly explained by the model on the junior and senior high school datasets. Therefore, it could be stated that although the model is suitable for the elementary school dataset, the model still needs some further adjustments to work with the junior and senior high school datasets.

For further conclusion, the model may have been suitable for the elementary school dataset because the education level system is more mature than the other two. Another point of view is that the results indicate that the elementary school educational development is more effective than the other two. It is also confirmed that according to clustering method itself, the data preprocessing gives a significant impact on the cluster creation. The impact will give influence to the analysis.

Based on the second finding, it is necessary to work with other parameters that have some potential influences to adjust the model in the near future.

References

[1] W. W. Leontief, Input-Output Economics, Science American, 185(4)(1951) 15–21.

[2] D. R. Entwisle and R. Conviser, Input-Output Analysis in Education, The High School Journal, 52(4)(1969) 192–198.

[3] N. S. Glasman and I. Biniaminov, Input-Output Analyses of Schools, Review of Educational Research, 51(4)(1981) 509–539.

[4] A. Olasunkanmi and M. Mabel, An input-output analysis of public and private secondary schools in Lagos, International Journal of Humanities and Social Science, 2(18)(2012) .

[5] C. Lei and S. Guicheng, Research of higher education input-output efficiency based on principal component analysis, MSIE 2011, (2011)174–177.

[6] H. G. Kim and N. Lee, Input-Output Analysis Of The Effect Of Education And Research ON The Korean Economic Structure, Journal of Applied Business Research (JABR), 14(2)(2011) 129–138.

[7] F. Wijayanto, Clustering Analysis on Indonesian Education Quality Performance Using Input-Output Model, Adv. Sci. Lett., (forthcoming).

[8] E. A. Licumba, J. Dzator, and J. X. Zhang, Gender Equality in Education and Economic Growth in Selected Southern African Countries., Journal of Developing Areas, 49(6)(2015) 349–360.

[9] F. Reza and T. Widodo, The Impact of Education on Economic Growth in Indonesia, Journal of Indonesian Economy and Business : JIEB., 28 Yogyakarta, (1)(2013) 23–44,159.

[10] Y. Azzizah, Socio-Economic Factors on Indonesia Education Disparity, International Education Studies, 8(12)(2015) 218.

[11] F. Wijayanto, Indonesia education quality: Does distance to the capital matter? (A clustering approach on elementary school intakes and outputs qualities), 2015 International Conference on Science and Technology (TICST), Pathum Tani, Thailand : IEEE, (2015)318–322.

[12] V. Dusen and G. C, The Virtual Campus: Technology and Reform in Higher Education, ASHE-ERIC Higher Education Report, 25(5)(2014) .

[13] V. Strauss, Nelson Mandela on the power of education, The Washington Post, (2013).

[14] K. P. D. Kebudayaan, Infografi Pendidikan Tahun 2011/2012, Jakarta, (2013).

[15] B. P. Statistik, Perkembangan Beberapa Indikator Utama Sosial-Ekonomi Indonesia, Jakarta, (2012).

[16] J. Scheerens, Process indicators of school functioning: A selection based on the research literature on school effectiveness, Studies in Educational Evaluation, 17(2)(1991) .

[17] N. Norwani, R. Yusof, and M. Abdullah, Learning Outcomes at Higher Education Institutions: To What Extent Do Institutional Environments Contribute?, The Journal of Human and Adult Learning, 5(1)(2009) .

[18] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, Top 10 algorithms in data mining, Knowledge and Information Systems, 14(1)(2008) 1–37.

[19] L. Korayem, M. Khorsid, and S. S. Kassem, A Hybrid K-Means Metaheuristic Algorithm to Solve a Class of Vehicle Routing Problems, Adv. Sci. Lett., 21(12)(2015) 3720–3722.

[20] M. Hedvicakova, A. Pozdilkova, P. K. Stranska, and L. Svobodova, Analysis of Mobile Social Networks Using Clustering, Adv. Sci. Lett., 22(5)(2016) 1273–1277.

[21] I. Dewi Tamara and S. Revina, Indonesian Mutual Funds Classification Using Clustering Method, Adv. Sci. Lett., 21(4)(2015) 826–829.

[22] R. Xu and D. Wunsch, Survey of clustering algorithms, IEEE Transactions on Neural Networks, 16(3)(2005) 645 – 678.

[23] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, San Francisco, (2006).

[24] D. T. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, (2005).

[25] M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs, NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set, Journal of Statistical Software, 61(6)(2014) 1–36.

[26] R Core Team, R: A Language and Environment for Statistical Computing, Vienna, Austria, (2013).