Hierarchical Regional Disparities and Potential Sector Identification Using Modified Agglomerative Clustering

T A Munandar¹, Azhari², A Mushdholifah³, and L Arsyad⁴

¹Doctoral Student of Computer Sc. and Electronic Dept., Math. And Natural Science Faculty, Universitas Gadjah Mada, Yogyakarta, INDONESIA
²Computer Sc. and Electronic Dept., Math. And Natural Science Faculty, Universitas Gadjah Mada, Yogyakarta, INDONESIA
³Economics Department, Faculty of Economics and Business, Universitas Gadjah Mada, Yogyakarta, INDONESIA
⁴Informatic Eng., Dept., Information Technology Faculty, Universitas Serang Raya

*tb.ai.m@mail.ugm.ac.id

Abstract. Disparities in regional development methods are commonly identified using the Klassen Typology and Location Quotient. Both methods typically use the data on the gross regional domestic product (GRDP) sectors of a particular region. The Klassen approach can identify regional disparities by classifying the GRDP sector data into four classes, namely Quadrants I, II, III, and IV. Each quadrant indicates a certain level of regional disparities based on the GRDP sector value of the said region. Meanwhile, the Location Quotient (LQ) is usually used to identify potential sectors in a particular region so as to determine which sectors are potential and which ones are not potential. LQ classifies each sector into three classes namely, the basic sector, the non-basic sector with a competitive advantage, and the non-basic sector which can only meet its own necessities. Both Klassen Typology and LQ are unable to visualize the relationship of achievements in the development clearly of each region and sector. This research aimed to develop a new approach to the identification of disparities in regional development in the form of hierarchical clustering. The method of Hierarchical Agglomerative Clustering (HAC) was employed as the basis of the hierarchical clustering model for identifying disparities in regional development. Modifications were made to HAC using the Klassen Typology and LQ. Then, HAC which had been modified using the Klassen Typology was called MHACK while HAC which had been modified using LQ was called MACLoQ. Both algorithms can be used to identify regional disparities (MHACK) and potential sectors (MACLoQ), respectively, in the form of hierarchical clusters. Based on the MHACK in 31 regencies in Central Java Province, it is identified that 3 regencies (Demak, Jepara, and Magelang City) fall into the category of developed and rapidly-growing regions, while the other 28 regencies fall into the category of developed but depressed regions. Results of the MACLoQ implementation suggest that there is only 1 regency which falls into the basic-sector category (Banyumas), while the other regencies fall into the non-basic non-competitive sector category.
1. Introduction
Disparities in the development sector are a multi-faceted issue with which many countries all over the world deal and are part of the development process itself [1]. Disparities in the development sector cannot be separated from the concept to determine potential sectors and they are commonly used in conjunction [2],[3],[4],[5],[6],[7],[8]. Disparities in the regional development are usually identified using the Klassen Typology, while potential sectors are commonly determined using Location Quotient (LQ). Both approaches cannot specifically visualize and categorize the GRDP sectors by the development achievement proximity between sectors, both within the region and between regions. Results of the analyses of these two approaches can only classify GRDP sectors according to the standard value interval. In fact, the GRDP sector value indicates the relationship between development achievements.

The inability of the Klassen Typology and LQ to show the relationship of such achievements poses problems to an effort to determine the direction of future development. This is unfortunate as by identifying which sectors have a good achievement level and which sectors with a quite similar achievement level, it will be easy to determine which sectors should be a priority or vice versa. This also applies for the achievements of regional development.

This research implemented the approach of Modified Hierarchical Agglomerative Clustering with Klassen (MHACK) to identify disparities in the regional development. Moreover, it also employed another approach, i.e. Modified Agglomerative Clustering with Location Quotient (MACLoQ), to determine the potential sectors owned by a particular region. Both approaches were resulted from modifying the algorithms of Hierarchical Agglomerative Clustering (HAC) with the methods of the Klassen Typology and Location Quotient (LQ). The combination between HAC and Klassen was referred to as MHACK while the combination between HAC and LQ was referred to as MACLoQ.

This paper is divided into five important sections. The first section is the introduction that describes the research background. The second section is the review of literatures on the problem to be solved. The third section presents the research methodology. The fourth section describes the results and discussion of the research undertaken. Lastly, the fifth section contains the research conclusions.

2. The Classification of Regional Disparities and Potential Sectors
Regional disparities are commonly classified using the Klassen Typology [2],[3],[4],[5]. The classification using this approach usually categorizes regions into four areas of disparities, namely Quadrants I, II, III, and IV. Quadrant I is comprised of developed and rapidly-growing regions; Quadrant II is comprised of developed but depressed regions; Quadrant III is comprised of potential regions or regions which can develop further; and Quadrant IV is comprised of relatively underdeveloped regions [6].

As for a region’s potential sectors, they are usually classified using the Location Quotient (LQ) method [7], [9]. This method classifies GRDP sectors into three categories based on the LQ value, they are the ‘basic sector’ category with an LQ value by > 1; the ‘non-basic non-competitive advantage sector’ category with an LQ value by = 1; and the ‘non-basic sector which can only meets its own necessities’ category with an LQ value by < 1. The resulting classifications of both Klassen and Location Quotient (LQ) are presented in Figures 1 and Figure 2, respectively.

![Figure 1. Disparities classification by Klassen](image1)

![Figure 2. Disparities classification by LQ](image2)
The Klassen approach is often used to identify disparities in regional development which is followed by identifying the potential of the region’s economic sector using LQ. However, both methods remain unable to optimally display the resulting classification information. The Klassen Typology remains unable to visualize development proximity according to the regional disparities being analyzed. Consequently, it is difficult to determine which regions should be a priority in the future development. On the other hand, LQ can only categorize potential regions based on the existing sectors, without the ability to see the proximity between potential sectors in a region. Therefore, it will be difficult to determine sectors which should be given priority in the future development.

Some other research used different approaches to classifying regional disparities based on economic indicators, such as the K-means clustering technique [10], [11], [12]. Besides K-means, regional disparities were also classified using the Hierarchical Agglomerative Clustering approach [13], [14], [15]. The resulting disparity classification using the clustering technique is able to visualize the proximity of disparities between regions. Nevertheless, it remains difficult to determine the exact number of clusters formed thus inhibiting deeper interpretations.

3. MHACK and MACLoQ Approaches

Modified Hierarchical Agglomerative Clustering with Klassen (MHACK) is a clustering method that combines the algorithms of Hierarchical Agglomerative Clustering (HAC) with the Klassen method. MHACK is developed to optimize the resulting classification of disparities in the regional development by showing the proximity of disparities amongst regions based on the clusters formed. MHACK algorithms proceed according to the following sequence:

a. Input the GRDP sector data of the regencies and the province (from two years or more)

b. Calculate the growth rate of the regencies and the province

c. Create the main cluster using the Klassen Typology

d. Calculate the distance of the data object in each main cluster, the resulting distance forms cluster singleton.

e. Determine the cluster set based on the cluster singleton in the main clusters

f. Combine two closest clusters

g. Update the distance data

h. Repeat steps e and f

Visualization of the resulting clusters using MHACK is shown in Figure 3.

![Figure 3. Disparities grouping with MHACK](image)

The work mechanism of the Modified Agglomerative Clustering with Location Quotient (MACLoW) is basically similar to that of the MHACK algorithms. The following are the classification stages of these algorithms:
a. Input the GRDP sector data of the regencies  
b. Calculate the LQ value of the GRDP sectors each regencies  
c. Create the main clusters based on the LQ value  
d. Calculate the distance of the LQ value of each main cluster, the resulting distance forms cluster singleton.  
e. Determine the cluster set based on the cluster singleton of the main clusters  
f. Combine two closest clusters  
g. Update the distance data  
h. Repeat steps e and f

Visualization of the resulting clusters using MACLoQ is shown in Figure 4.

Figure 4. Potential sectors grouping with MACLoQ

4. Research Methodology  
This research began with collecting the GRDP sector data for Central Java Province. The data were obtained online from the Office of the Statistics Indonesia. The data on the GRDP sectors in 2012 of 31 regencies situated in Central Java Province were used in this research. A total of 297 datasets were used to classify disparities in the regional development using MHACK and to identify their potential sectors using MACLoQ. Afterwards, the classification results were interpreted in order to determine the existing disparities as well as the potential sectors in a particular region.

5. Results and Discussion  
The implementation of MHACK upon the GRDP sector dataset of 31 regencies in Central Java Province shows that this province is classified into two main clusters. These two main clusters display disparities in Central Java Province. The first cluster (Quadrant I) consists of Demak Regency (1), Jepara Regency (2), and Magelang City (3). The first cluster shows that the regions belonging to this category are those developed and rapidly-growing regions. The other twenty-eight regencies fall into the second main cluster (Quadrant II), i.e. the category of developed but depressed regions. The graphic visualization of the first cluster (Quadrant I) is shown in Figure 5.
Figure 5. Cluster visualization by MHACK for Quadrant II

Figure 6. Cluster visualization by MACLoQ for LQ < 1 with the number of Trees shown amounts to T = 30

Figure 5 shows that Demak (1) and Jepara (2) Regencies have the proximity of development achievements. Both regencies were then clustered to form a new group. The Demak-Jepara group was then clustered again hierarchically with Magelang City (3) due to the proximity of development achievements. This resulting cluster shows the hierarchy of development achievements that cannot be done by merely using the Klassen method. Moreover, the resulting clustering using MHACK can visually help policy makers to determine which regions will be a priority in the future development. Further analyses were performed using results of the clustering using MACLoQ to determine the potential sectors owned by the regions belonging to the first main cluster based on the classification process using MHACK. There are nine GRDP sectors to be identified, namely agriculture, livestock, forestry, fisheries (S1); mining and quarrying (S2); manufacturing industry (S3); electricity, gas, and clean water (S4); construction (S5); trade, hotels, and restaurants (S6); transport and communication (S7); finance, real estate, and business services (S8); and services (S9).

Based on the clustering results using MACLoQ, it is revealed that the GRDP sectors of three regencies (Demak Regency, Jepara Regency, and Magelang City) fall into the category of non-basic sectors which can only meet their own necessities. The clustering results using MACLoQ also provide the information about the proximity of the GRDP sectors of both within one region and between region. Tables 1, 2, and 3 show the proximity relationship of GRDP sectors for Demak Regency, Jepara Regency, and Magelang City, respectively.

| No. | Sector Id | LQ value | Nearest sectors and regency (NS) | LQ value for NS |
|-----|-----------|----------|---------------------------------|----------------|
| 1   | S1        | 0.003469477 | S9 – Kab. Brebes              | 0.003422183    |
| 2   | S2        | 0.000235471  | S3 – Kab. Banjarnegara       | 0.000239364    |
| 3   | S3        | 0.00041489   | S4 – Kab. Purbalingga        | 0.000404669    |
| 4   | S4        | 0.001451013  | S9 – Kab. Klaten             | 0.001479407    |
| 5   | S5        | 0.001211444  | S1 – Kab. Rembang            | 0.00121303     |
Table 1. Cont.

| No | Sector Id | LQ value   | Nearest sectors and regency (NS)      | LQ value for NS |
|----|-----------|------------|--------------------------------------|----------------|
| 6  | S6        | 0,001135533| S1 – Kab. Sukoharjo                   | 0,001138444    |
| 7  | S7        | 0,001239627| S3 – Kab. Cilacap                     | 0,001250645    |
|    |           |            | S4 – Kab. Kudus                       | 0,001251021    |
|    |           |            | S9 – Kab. Banjarnegara                | 0,001245528    |
| 8  | S8        | 0,001590671| S1 – Kab. Kendal                      | 0,001589442    |
| 9  | S9        | 0,001579449| S3 – Kab. Brebes                      | 0,0015747      |

Table 2. Proximity relationship of GRDP sectors for Jepara Regency

| No | Sector Id | LQ value   | Nearest sectors and regency (NS)    | LQ value for NS |
|----|-----------|------------|-------------------------------------|----------------|
| 1  | S1        | 0,002595445| S4 – Kota Surakarta                  | 0,003053847    |
|    |           |            | S8 – Kota Surakarta                  | 0,003028373    |
|    |           |            | S5 – Kab. Wonogiri                   | 0,002984049    |
| 2  | S2        | 0,001285988| S4 – Kab. Semarang                   | 0,001277364    |
| 3  | S3        | 0,001752137| S2 – Kab. Tegal                      | 0,001812443    |
| 4  | S4        | 0,00294148 | S5 – Kab. Brebes                     | 0,002920036    |
| 5  | S5        | 0,002380291| S3 – Kab. Boyolali                   | 0,00237865     |
| 6  | S6        | 0,002004458| S6 – Kab. Cilacap                    | 0,002003477    |
| 7  | S7        | 0,002120167| S6 – Kab. Cilacap                    | 0,002003477    |
|    |           |            | S6 – Kab. Jepara                     | 0,002004458    |
|    |           |            | S7 – Kota Surakarta                  | 0,002059055    |
| 8  | S8        | 0,003910257| S1 – Kab. Demak                      | 0,003469477    |
|    |           |            | S9 – Kab. Brebes                     | 0,003422183    |
|    |           |            | S6 – Kab. Tegal                      | 0,003352158    |
|    |           |            | S5 – Kab. Boyolali                   | 0,003250718    |
|    |           |            | S7 – Kab. Boyolali                   | 0,003722701    |
| 9  | S9        | 0,002202596| S7 – Kab. Cilacap                    | 0,002250237    |
|    |           |            | S2 – Kab. Magelang                   | 0,002279577    |

Table 3. Proximity relationship of GRDP sectors for City of Magelang

| No | Sector Id | LQ value   | Nearest sectors and regency (NS)    | LQ value for NS |
|----|-----------|------------|-------------------------------------|----------------|
| 1  | S1        | 0,002595445| S4 – Kota Surakarta                  | 0,003053847    |
|    |           |            | S8 – Kota Surakarta                  | 0,003028373    |
|    |           |            | S5 – Kab. Wonogiri                   | 0,002984049    |
| 2  | S2        | 0,001285988| S4 – Kab. Semarang                   | 0,001277364    |
| 3  | S3        | 0,001752137| S2 – Kab. Tegal                      | 0,001812443    |
| 4  | S4        | 0,00294148 | S5 – Kab. Brebes                     | 0,002920036    |
| 5  | S5        | 0,002380291| S3 – Kab. Boyolali                   | 0,00237865     |
| 6  | S6        | 0,002004458| S6 – Kab. Cilacap                    | 0,002003477    |
| 7  | S7        | 0,002120167| S6 – Kab. Cilacap                    | 0,002003477    |
|    |           |            | S6 – Kab. Jepara                     | 0,002004458    |
|    |           |            | S7 – Kota Surakarta                  | 0,002059055    |
| 8  | S8        | 0,003910257| S1 – Kab. Demak                      | 0,003469477    |
|    |           |            | S9 – Kab. Brebes                     | 0,003422183    |
|    |           |            | S6 – Kab. Tegal                      | 0,003352158    |
|    |           |            | S5 – Kab. Boyolali                   | 0,003250718    |
|    |           |            | S7 – Kab. Boyolali                   | 0,003722701    |
| 9  | S9        | 0,002202596| S7 – Kab. Cilacap                    | 0,002250237    |
|    |           |            | S2 – Kab. Magelang                   | 0,002279577    |
Table 3. Cont.

|   |   |   |   |
|---|---|---|---|
| 6 | S6 | 8,4207E-05 | S3 – Kab. Blora | 8,48244E-05 |
| 7 | S7 | 0.000834604 | S1 – Kab. Pekalongan | 0.00079835 |
|   |   |   | S9 – Kab. Pemalang | 0.000818913 |
| 8 | S8 | 0.000699587 | S4 – Kota Magelang | 0.000703473 |
|   |   |   | S5 – Kab. Banjarnegera | 0.000701741 |
| 9 | S9 | 0.000890683 | S8 – Kab. Purbalingga | 0.000892876 |

Tables 1 to 3 show the proximity between the potential sectors of Demak Regency, Jepara Regency, and Magelang City and the other sectors for both sectors within a region and sectors between regions. Figure 6 illustrates cluster visualization using MACLoQ with the number of Trees shown amounts to \( T = 30 \).

The cluster output consistency test was undertaken by calculating the value of the cophenet distance to examine whether the clusters formed have been optimal or not. If the cophenet value is close to 1, it means that the clusters formed are very good. The cophenet values for clustering using MHACK is 0.9950 for the outcome of the first main cluster and 0.9154 for the outcome of the first main cluster. As for the clustering using MACLoQ, the cophenet values are equal to 0.8203 and 0.9024 for the first cluster (non-basic, competitive) and the third cluster (non-basic, non-competitive), respectively. Thus, it can be concluded that the output of the clustering using both MHACK and MACLoQ is very good.

6. Conclusions

The research conducted indicates that MHACK algorithms can be used as a new approach to identifying disparities in regional development between regions. In addition, it can also show the proximity relationship between development achievements between regions of the same disparity group, things which the conventional Klassen method which is currently used widely cannot do. Meanwhile, the MACLoQ algorithms can also identify the proximity of potential sectors both within a region and between regions, in addition to classifying the potential sectors owned by the said region. Results of the cluster output consistency test indicate that the clusters formed are deemed very good with a cophenet value close to 1. Results of clustering using MHACK can be used to assist policy makers in determining which regions which should get priority in the development sector, while the results of clustering using MACLoQ can be used to determine which sectors in an region to be developed which should get priority.

References

[1] Williamson, J.G., 1965, *Regional Inequality and the Process of National Development: A Description of the Patterns*, Source: Economic Development and Cultural Change, Vol. 13, No. 4, Part 2 (Jul., 1965), pp. 184

[2] Hariyanti, D. and Utha, M.A., 2016, Analysis of Determinants Sectors Regional Development at 33 Provinces in Indonesia, *International Journal of Sustainable Development*, 09, (03)

[3] Suwandi, N., 2015, Economic Inequality among Districts in Keerom, Papua, Indonesia, *Journal of Social and Development Sciences*, Vol. 6, No. 3, pp. 30-34

[4] Endaryanto, T., Firdaus, M., Siregar, H., and Hakim, D.B., 2015, The Impact of Regional Expansion on Economic Structure: A Case Study in Lampung Province, Indonesia, *International Journal of Sciences: Basic and Applied Research (IJSBAR)* (2015) Volume 23, No 2, pp 1-18

[5] Fattah, S., and Rahman, A., 2013, Analysis of Regional Economic Development in the Regency/Municipality at South Sulawesi Province In Indonesia, *Journal of Economics and Sustainable Development*, Vol.4, No.1

[6] Karsinah, Putri, P.I., Rahayu, K.N., and Panjiputri, A.F., 2016, The Profile of Pekalongan as a
Center of Economic Growth at Tangkallangka Strategic Areas, *International Journal of Economics and Financial Issues*, Vol 6, Special Issue (S6)

[7] Sinaga, D., 2015, Determination Analysis of Leading Economic Sector Against Forming Region GDP in Simalungun, *International Journal of Innovative Research in Management*, issue 4 volume 3

[8] Sinurat, M., 2016, Cianjur Regency Regional Economy Potential Analysis, *International Journal Of Social Sciences*, Vol.45. No.1

[9] Bakaric, I.R. 2005. Uncovering Regional Disparities – the Use of Factor and Cluster Analysis. *Economic Trends and Economic Policy*. No. 105 , pp. 52-77

[10] Lukovics, Miklós. 2009. Measuring Regional Disparities on Competitiveness Basis. *JATEPress*, Szeged, pp. 39-53

[11] Soares, J.O., Marques, M.M.L dan Monteiro, C.M.F. 2003. A Multivariate Methodology To Uncover Regional Disparities: A Contribution To Improve European Union And Governmental Decisions. *European Journal of Operational Research* 145 (2003) 121–135.

[12] Kronthaler, F., 2003. A Study of the Competitiveness of Regions based on a Cluster Analysis: The Example of East Germany. *Institute for Economic Research Halle* (IWH)

[13] Vydrová, H. V., dan Novotná, Z., 2012. Evaluation Of Disparities In Living Standards Of Regions Of The Czech Republic. *Acta Universitatis Agriculturae Et Silviculturae Mendelianae Brunensis*. Volume LX 42 Number 4

[14] Nosova, O., 2013. The Innovation Development in Ukraine: Problems and Development Perspectives. *Int., J., Of Innovation And Business Strategy*. Vol. 02/August

[15] Poledníková, E., 2014. Comparing Regions’ Ranking by MCDM methods: the Case of Visegrad Countries. *WSEAS TRANSACTIONS on BUSINESS and ECONOMICS*, Volume 11, 2014, pp. 496 – 507