Provincial economic level analysis in Indonesia based on the geothermal energy potential and growth regional domestic products using cluster analysis

G M Idroes¹, S Syahnur¹, S A. Majid¹, N R Sasmita², R Idroes³
¹Department of Economics, Faculty of Economics and Business, University of Syiah Kuala, Banda Aceh 23111, Indonesia
²Computational and Applied Statistics Research Group, Department of Statistics, Faculty of Mathematics and Natural Sciences, University of Syiah Kuala, Banda Aceh 23111, Indonesia
³Department of Chemistry, Department of Statistics, Faculty of Mathematics and Natural Sciences, University of Syiah Kuala, Banda Aceh 23111, Indonesia

*E-mail: kabari_sofyan@unsyiah.ac.id

Abstract. This study aims to determine the provincial economic level in Indonesia based on geothermal energy potential and growing regional domestic product (GRDP). The data used in this research is secondary data. The type of data is a cross-section from 34 Provinces in Indonesia in 2018. Geothermal Energy Potential Data is obtained from the Ministry of Energy and Mineral Resources (ESDM), and GRDP data for each province is obtained from the Indonesian Central Statistics Agency (BPS). This study divided data into 5 clusters using the k-means method. The results showed that there are 2 provinces in cluster 1, 1 province in cluster 2, 8 provinces in cluster 3, 22 provinces in cluster 4, and 1 province in cluster 5.

1. Introduction

Indonesia has the greatest potential for geothermal energy in the world with a total of 28.91 GW because it is located in a ring of fire in a volcanic route that is spread over 312 locations on several islands such as Sumatra, Java, Bali, Nusa Tenggara and Sulawesi [1]. Geothermal energy is environmentally friendly and quite economical [2] when compared to fossil fuels.

The resource of geothermal potential in Indonesia is predicted by about 40% of the total of the world's geothermal reserves of around 28,617 MW. Whereas, only about 4.5% is being exploited as electrical energy in the country. The power plant comes from a minimum of 10 locations: Darajat (260 MWe), Dieng (60 MWe), Kamojang (200 MWe), Gunung Salak (377 MWe), Sibayak (11 MWe), Lahendong (87 MWe), Wayang Windu (227 MWe), Ulu Belu – South Sumatra (110 MWe), Ulumbu – Flores (5 MWe) and Mataloko (2.5 MWe) [3].

Economic growth is a process of changing a country's economic conditions on an ongoing basis towards better conditions for a certain period. The economic growth of a region can be measured with the Gross Regional Domestic Product (GRDP). The existence of economic growth can be regarded as a success in heading for a better economic level.
This study aims to determine the economic level by clustering each province in Indonesia based on the potential of geothermal energy and the GRDP.

2. Literature Review

2.1 GRDP
GRDP (Gross Regional Domestic Product) is the total value added of goods and services resulting from all economic activities in all regions in a certain year period which is generally within one year.

2.2 Geothermal Potential
Geothermal potential in this study is divided into 5 potentials, namely speculative, hypothetic, possible, probable, and proven.
1) Speculative resources: the estimation of geothermal potential based on the existence of geothermal manifestations and heat flow.
2) Hypothetical resources: determined from regional geologic surveys and geochemistry of thermal features. Stored heat calculations and estimates used for resource sizing.
3) Possible reserves: the potential for geothermal energy in the earth is estimated using integrated surface geoscientific survey data.
4) Probable reserves: the estimation of potential geothermal subsurface by using integrated surface geoscientific survey data and the result of one discovery well.
5) Proven reserves: the potential geothermal energy below the surface is estimated using a geoscience work survey that includes one well and illustrates two production wells to obtain data on the subsurface of the system [4].

2.3 Hierarchy Method
The hierarchy method is a grouping method where the results are presented on several levels or in stages from groups n, (n-1), ..., 1 [5]. The commonly used distance function is the Euclid distance, which is defined as the distance between the i-th and k-th observations. Euclid's distance formula from the i-th object to the h-th object is formulated in equation (1):

\[ d(i, h) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{hk})^2} \]

Some kind of hierarchical method of merging (agglomerative) based on linkage includes single linkage, complete linkage, average environment and ward's method [3].

2.4 Non Hierarchy Method
In contrast to the hierarchical method, the non-hierarchical grouping procedure is not done in stages and the number of groups is also determined in advance. Methods that are classified as non-hierarchical grouping include k-means.

Grouping using the k-means method is based on the value of the membership function. The membership function is based on the minimum distance between objects and the center of the cluster (centroid). The k-means algorithm aims to minimize the objective function which is a quadratic error function. For example, there are n objects and p variables. The distance between the i-th object and the l-group is calculated using Euclid's distance squared formulated according to equation (2):

\[ D(i, l) = \sum_{j=1}^{p} [X(i, j) - \bar{x}(l, j)]^2 \]
With,

\(X(i,j)\): value between the i-th object with the j-th variable; \(X(l,j)\): the average of the j-th variable against the l-th group; \(D[i,l(i)]\): Euclid's distance between the i-th object and the cluster average of the object cluster (centroid) [5].

3. Research Method

3.1. Data Sources and Types
The data used in this research is secondary data. The entire data is in the form of cross-section data (cross-border data) from 34 Provinces in Indonesia in 2018. Geothermal Energy Potential Data is obtained from the Ministry of Energy and Mineral Resources (ESDM) and GRDP data from 34 Provinces in Indonesia is obtained from the Central Statistics Agency (BPS) Indonesia.

3.2. Data Collection Technique
The data of this research were collected in the same year in the form of tabulation by conducting the data cleaning process. Then, the data is analyzed.

3.3. Data Analysis Technique
Data analysis techniques used in this study are descriptive and cluster analysis, which is by describing the data that has been collected properly without and to provide conclusions that are generally accepted and group them according to the size of the similarity between objects using the K-Means method.

Descriptive analysis is in the form of a summary of the data used for each variable, such as the size of the central tendency and the size of the spread [6]. Cluster analysis used in this study follows the flowchart shown in Figure.

![Flowchart Algoritma K-Means Clustering](image)

**Figure 1.** Flowchart Algoritma K-Means Clustering

3.4. Software
The software used in this study is Microsoft EXCEL® and IBM SPSS® Statistics version 22.

4. Results and Discussion

4.1. Descriptive Analysis
Statistically, a description of the economic level for each province in Indonesia can be determined based on the size of concentration and the size of its distribution. By using this measure of concentration, the average level of the economy in Indonesia will be known.
Table 1. Descriptive Statistics of 2017 Geothermal energy potential and GRDP 2017

| Variable      | Minimum | Maximum | Average | Standard Deviation |
|---------------|---------|---------|---------|--------------------|
| Speculative   | 0       | 1125    | 188.97  | 226.898            |
| Hypothetic    | 0       | 734     | 118.74  | 187.247            |
| Possible      | 0       | 1583    | 377.03  | 459.635            |
| Probable      | 0       | 1283    | 72.26   | 223.582            |
| Proven        | 0       | 1550    | 87.12   | 273.111            |
| GRDP          | -4.56   | 7.92    | 5.27    | 2.06               |

In Table 1 it can be seen that all geothermal potentials have a minimum value of 0 Mwe, while the maximum potential for geothermal energy is at an estimated potential of 1583 Mwe, after that the estimated potential of 377.03 Mwe achieves the highest average geothermal energy potential, then the standard deviation energy value geothermal potential is in the Suspected potential of 459.63 Mwe.

In Table 1 it can be seen that the minimum GRDP value is -4.56%, and the maximum GRDP value is 7.92%, while the average GRDP value is 5.27%, and the highest standard deviation GRDP value is 2.06%.

4.2. Cluster Analysis

Using the K-mean Method as a method in non-hierarchy to classify each province in Indonesia based on geothermal energy potential and GRDP variables for each province, it is necessary to standardize values. In detail, the results of the initial standardization are listed in Table 2.

Table 2. Initial grouping center on standardized values

| Variable          | Cluster | 1     | 2     | 3     | 4     | 5     |
|-------------------|---------|-------|-------|-------|-------|-------|
| Zscore(Speculative) | -.27312 | 4.12533 | .15879 | -.48467 | -.83284 |
| Zscore(Hypothetic)  | 1.37394 | 3.28585 | 2.79986 | -.55400 | -.63411 |
| Zscore(Possible)    | 2.36486 | 2.62376 | 1.15302 | -.82028 | -.82028 |
| Zscore(Probable)    | -.32321 | 5.41516 | .25823  | -.32321 | -.32321 |
| Zscore(Proven)      | -.31898 | 5.35637 | .45726  | -.31898 | -.31898 |
| Zscore(GRDP)        | 2.82840 | 2.50349 | -.02503 | -.63797 | 3.21729 |

The standardization process carried out refers to the z-score with the following conditions: A negative value (-) means the data is below the total average and a positive value (+) means the data is above the total average.

Table 2 shows the initial data of the clustering process before iterating the cluster analysis. In the grouping process in this study, the number of iterations carried out to get the most optimal results requires an iteration of 5 times. Then, Table 3 shows the final grouping center on the standardized value.

Table 3. The final grouping center is the standardized value

| Variable          | Cluster | 1     | 2     | 3     | 4     | 5     |
|-------------------|---------|-------|-------|-------|-------|-------|
| Zscore(Speculative) | -.34364 | 4.12533 | .68271 | -.36668 | -.83284 |
| Zscore(Hypothetic)  | 1.03748 | 3.28585 | .85190 | -.52463 | -.63411 |
| Zscore(Possible)    | 1.82095 | 2.62376 | .82450 | -.54734 | -.82028 |
| Zscore(Probable)    | -.12194 | 5.41516 | .15759  | -.27767 | -.32321 |
| Zscore(Proven)      | .24855  | 3.35637 | .10758  | -.29069 | -.31898 |
| Zscore(GRDP)        | 2.12649 | 2.50349 | -.25074 | -.36217 | 3.21729 |
Furthermore, Table 4 explains the differences in the variables formed in this cluster can be seen from the F value and the probability value for each variable. The greater the value of F and significantly smaller than 0.05, the greater the difference in variables in the cluster formed.

Table 4. ANOVA in cluster analysis

| Variable         | Cluster | Error | F Mean Square | Sig. df |
|------------------|---------|-------|---------------|---------|
|                  |         | Mean Square | df | Mean Square | df |
| Zscore(Speculative) | 6.159   | 4 | .288 | 29 | 21.351 | .000 |
| Zscore(Hypothetic)  | 6.303   | 4 | .269 | 29 | 23.473 | .000 |
| Zscore(Possible)   | 6.554   | 4 | .234 | 29 | 28.026 | .000 |
| Zscore(Probable)   | 7.838   | 4 | .057 | 29 | 138.026 | .000 |
| Zscore(Proven)     | 7.717   | 4 | .074 | 29 | 104.949 | .000 |

Thus the results of the clusters obtained in this study indicate that the variable probably most shows the differences between the provinces in the five clusters formed. This can be indicated by the value of F Maybe = 138,026 and significant = 0.00 for other variables interpreted with the same thing.

Table 5. Cluster Analysis Results Based on Geothermal Potential Indicators and GRDP

| No | Cluster | Provinces |
|----|---------|-----------|
| 1  | Cluster 1 | (1) Jawa Tengah, (2) Jawa Timur |
| 2  | Cluster 2 | (1) Jawa Barat |
| 3  | Cluster 3 | (1) Aceh, (2) Sumatera Utara, (3) Sumatera Barat, (4) Jambi, (5) Sumatera Selatan, (5) Bengkulu, (6) Lampung, (7) Nusa tenggara Timur |
| 4  | Cluster 4 | (1) Riau, (2) Kepulauan Bangka belitung, (3) Kepulauan Riau, (4) DI Yogyakarta, (5) Banten, (6) Bali, (7) Nusa tenggara barat, (8) Kalimantan barat, (9) Kalimantan tengah, (10) Kalimantan selatan, (11) Kalimantan timur, (12) Kalimantan utara, (13) Sulawesi utara, (14) Sulawesi tengah, (15) Sulawesi selatan, (16) Sulawesi tenggara, (17) Gorontalo, (18) Sulawesi barat, (19) Maluku, (20) Maluku utara, (21) Papua barat, (22) Papua |
| 5  | Cluster 5 | (1) DKI Jakarta |

Figure 2. Visualization of the results of the cluster
Based on the results of cluster analysis, as shown in Table 5, it is found that there are 2 provinces in cluster 1, 1 province is in cluster 2, 8 provinces are in cluster 3, 22 provinces are in cluster 4 and 1 province is in cluster 5. In detail, cluster 1 is a province with potential low speculative and proven and high potential, hypothetical, probable, and high GRDP. In cluster 2 it contains provinces with high speculative potential, hypotheses, Possible, Probable, proven, and GRDP. In cluster 3 it contains provinces with high speculative, hypothetical, Possible, Probable, and proven potential and low GRDP. In cluster 4 it contains provinces with low speculative potential, hypotheses, possible, Probable, proven, and GRDP. In cluster 5 it contains provinces with low speculative, hypothetical, probable, and proven potential and high GRDP. Figure 2 illustrates the results.

5. Conclusions
The results of descriptive statistical analysis for geothermal potential variables indicate that the Estimated Potential is the potential with the lowest minimum value, the highest maximum value, the highest average value and the highest standard deviation value. Then, the minimum GRDP value is -4.56%, and the maximum GRDP value is 7.92%, while for the average GRDP value is 5.27%, and the highest standard deviation GRDP value is 2.06%.

Cluster analysis results from 34 provinces based on geothermal potential and GRDP can be grouped into 5 groups. The results of the grouping are classified into several regions; cluster 1 is a province with low speculative and proven potential and high hypothetical, Possible, Probable, and high GRDP potential, namely: Central Java and East Java. Cluster 2 is a province with high speculative, hypothetical, Possible, probable, and high GRDP potential, is West Java. Then, cluster 3 is a province with high speculative, hypothetical, Possible, probable and low proven provinces, namely Aceh, North Sumatra, West Sumatra, Jambi, South Sumatra, Bengkulu, Bengkulu, Lampung, and East Southeast Nusa Tenggara. Furthermore, cluster 4 is a province with a low speculative, hypothetical, probable, probable, and low GRDP potential, namely Riau, kep. Pacific Islands, Kep. Riau, in Yogyakarta, Banten, Bali, West Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, North Sulawesi, Central Sulawesi, South Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, North Maluku West Papua, and Papua. Finally, cluster 5 is a province with low speculative, hypothetical, Possible, Probable, and proven potential, namely DKI Jakarta.

References
[1] Pambudi N A 2018 Renew. Sustain. Energy Rev. 81 2893–901.
[2] Idroes R, Yusuf M, Alatas M, Subhan, Lala A, Muslem, Suhendra R, Idroes G M, Suhendrayatna, Marwan, Riza M 2019 IOP Conf. Ser. Mater. Sci. Eng. 523 012010.
[3] Kuncoro M 2003 Metode riset untuk bisnis dan ekonomi. Jakarta: Erlangga. (in Bahasa Indonesia).
[4] Pertamina 2005 Pertamina Geothermal Development: Resources & Utilization: Publication of Pertamina Geothermal Division. Jakarta: Pertamina.
[5] Lazulfa I 2013 Analisis Cluster Kabupaten/Kota di Jawa Timur Berdasarkan Tingkat Pencemaran Udara. Doctoral Dissertation. Surabaya: Institut Teknologi Sepuluh Nopember. (in Bahasa Indonesia).
[6] Kristanto N H, Christopher A, Budi H. 2016 Jurnal Teknik Informatika dan Sistem Informasi. 2 pp. 9–15. (in Bahasa Indonesia).