Clustering the basic human category indicator levels in Banten region

Y Asri\(^1\), P Palupiningsih\(^1\), Haryono\(^2\), and R M Reviansya\(^1\)

1 Majoring Informatics Engineering, Sekolah Tinggi Teknik PLN, Jl. Lingkar Luar Duri Kosambi, Cengkareng, Jakarta Barat, Indonesia, 11750
2 Majoring Informatics Engineering, STMIK Bani Saleh Bekasi, Jl. Jl. Mayor M. Hasibuan No. 68, Margahayu, Bekasi Timur, Kota Bekasi, Jawa Barat, Indonesia 17113

Abstract — Indonesia has an inequality in human development, education provides the highest inequality. Loss that occurs in the education category reaches 20.8 percent. In the categories of decent living standards reaches 17.3 percent, and in the health category 16.5 percent. Most volunteers and the society take responsive action after getting information from the media or from the local communities. Therefore, this study aims to find out which areas have problems on indicators of basic human categories in a certain year. Regions will be grouped into three cluster groups, regions with low values, regions with medium values, and regions with high values. K-Means is an algorithm used to classify regions. The results of this study are, knowing the areas that have the lowest, medium, and highest values based on indicators of basic human categories in 2013 to 2017. Pandeglang and Serang Regencies are the lowest indicators such as life expectancy with a center point of 62.928 in 2013 and up to 64.03 in 2017, mean years of school is 6.299 to 6.666, per capita expenditure is 8411.72 to 9065.33. The cities of South Jakarta, East Jakarta, and Central Jakarta, areas that are at the highest indicators such as life expectancy with a center point of 72.938 to 73.107, expected years of school is 12.673 to 13.438, mean years of school is 10.469 to 10.851, and per capita expenditure with a center point of 17942.57 up to 18856.4.

1. Introduction
In terms of status, human development in DKI Jakarta Province has entered the status of "Very High" HDI (Human Development Index). This status is an increase compared to the previous year. While on average, Jakarta Capital City's HDI (Human Development Index) in 2010-2017 recorded a growth of 0.69 percent per year. All components that form the HDI (Human Development Index) also experience an increase. Age of life expectancy at birth in 2017 is 72.55 years, an increase compared to 2016 of 72.49 years. The year-long school year expectation reached 12.86 years, an increase from last year's 12.73 years. The average length of schooling increased from 10.88 years in the previous year to 11.02 years. The last component, namely per capita expenditure also increased from 17.5 million rupiah to 17.7 million rupiah\(^1\). Banten continues to increase. This can be seen from the increase in HDI (Human Development Index) numbers consistently during the period 2010-2016. Banten's HDI (Human Development Index) figure in 2016 reached 70.96. Means, the level of achievement of human development can be said to be around 71 percent of ideal human development conditions (ideal HDI = 100). However, its human development status has not experienced an increase, which is still in the "High" category (70 M HDI <80) \(^2\). Inequality in human development in Indonesia occurs in all
categories. Education provides an alarming portrait of inequality. Loss that occurs in the education category reaches 20.8 percent. In the standard of decent living standard, inequality reached 17.3 percent. While in the health category, the lowest inequality is recorded among other categories, namely 16.5 percent [3].

For this reason, an appropriate development model and strategy is needed, which can help improve the welfare of the community. There are factors that make the development index low, the factors are the basic categories of human beings who have indicators of life expectancy, long school expectations, average length of schooling, and per capita expenditure. Researchers have category indicator data that can be processed using clustering and K-Means algorithms. Researchers used a case study of Banten Province and DKI Jakarta, because Banten Province was a neighboring province of DKI, which would certainly affect all values of the category indicator. The purpose of this research is to be able to analyze the extent of the growth of each category indicator every year, each indicator will also be clustered to find out which regions have the smallest value in a group based on categorical indicators.

K-Means provides relatively good results when groups in different datasets are well separated. K-Means are also more efficient in computing time complexity. K-Means has several disadvantages related to forms and clusters in the dataset. For example, K-Means allows unsuccessful when searching for clusters that are overlapping, and also K-Means allows failure when data is not normalized [4].

2. Related Work

Research titled "Prediction of Chemotherapeutic Response in Bladder Cancer Using K-means Clustering of Dynamic Contrast-Enhanced (DCE) - MRI Pharmacokinetic Parameters". To implement K-Means grouping two pharmacokinetic parameters derived from 3T contrast Dynamic enhanced magnetic resonance imaging (DCE-MRI) for predicted chemotherapy response to bladder cancer at the time of cycle time middle. Clustering of each bladder tumor is partitioned be: cluster 1 (low and low Amp), cluster 2 (low and low High amp), cluster 3 (high and low Amp). Change from third VF clusters are found to be associated with the response of bladder tumors to chemotherapy. VF changes from cluster 2 are presented with the highest area value under the curve (0.96) and the highest sensitivity/specificity / accuracy (96% 100% / 97%) with selected cutting value (Huyen, et al., 2014).

The above research has in common that is using the K-Means Clustering method. Research differences exist in the cases faced. The second study, entitled "A K-Means-based Network Partition The Network Placement in Software Defined Controller Algorithm ". Software Defined Networking (SDN), a new paradigm of decoupling control logic from the forwarding package device, has attracted a lot of good attention from academics and industry. Because of the latency between the controller and the switch is an important factor for SDN, choosing the appropriate location for the controller to shorten latency is a big challenge. In this paper, they investigated the problem of multi-controller placement from latency minimization perspective. Different from the previous work, partition technique the network was introduced to simplify the problem. In particular, network partition problems and first time controller placement problems formulated. The optimized K-Means algorithm is then proposed to solve the problem. Extensive simulations are carried out and the results indicate that the proposed algorithm can greatly reduce latency; the maximum between their centroid and node is compared to the K- standard Means. Specifically, the maximum latency can reach 2,437 times more short of the average latency achieved by the K-Means standard (Wang, Zhao, Huang, Duan Li, 2016).

The research has in common that is using the K-Means Clustering method to solve a problem. The research difference is in the case faced, namely optimizing the K-Means algorithm to minimize latency, while the researchers only use the K-Means algorithm standard. The third study, entitled "Mapping of Teacher Spread in Provinces Banten by Using Spatial Clustering K-Means Method (Study case: Banten Province Region "). The level of education in Indonesia is still classified as low, one of the biggest contributors to education is the teacher's role. Teacher equity in Indonesia is not evenly distributed in all region, most teachers only occupy urban areas.
Based on data from the Central Bureau of Statistics in 2013, Banten Province has the lowest Human Development Index on the island of Java. Therefore, this study aims to map the distribution of teachers in the province of Banten. In this study grouping using the K-Means algorithm based on the number of teachers, number of students and number of schools at the level Primary School, Junior High School and School education Upper middle. Grouping results are then mapped with using Geographic Information Systems. The results of this study are mapping of Banten Province based on the level of education have short comings, sufficiency and excess teacher in accordance with district / city. The benefits of the results of this study are suggestions for the Banten Provincial Education Office in terms of teacher equity (Priambodo Prasetyo, 2018).

The above research has in common that is using the K-Means Clustering method to solve a problem. The above research has a relationship in the discussion of the problem, one of the basic categories of humans is education, which is certainly influenced by the quality of the teacher. Research differences are found in the cases faced, of course the data processed is different. The fourth study, entitled "The Influence of Income Inequality, Per capita Income, and Government Expenditures in the Health Sector Against the Health Sector in Indonesia "explains that human resources are one of the important capital in the development of a nation. One of the important aspects that influences human resources is the level of public health, where the health sector has a role important. A person's health status is the result of the interaction of various factors, namely internal and external factors. Internal factors consist of physical factors and psychology, while external factors consist of economic, education, environment and culture. This study aims to test and analyze the influenceincome inequality as measured by the Gini Ratio of the sector health measured by life expectancy in Indonesia in 2005 - 2013. In this study the regression equation uses a data panel with the Random Effect Model approach. The results of this study are income inequality, per capita income, and Government expenditure on health effects simultaneously on the health sector at Indonesia in 2005-2013 and is partial, inequality income, per capita income, and government expenditure in the sector health impacts on the health sector in Indonesia in 2005-2013(Hasanah & Ahmadi, 2017).

The above research has in common that is discussing the problem of one indicator of human categories. Research differences exist in the method approach.

3. Proposed Method
Research methodology is a scientific process or method to obtain data that will be used for research purposes.

3.1 Data Collection
The method used by researchers in data collection is in the form of interviews and observations. The basic human category indicator data can be obtained free of charge and can be accessed by the public on the website https://bps.go.id. Completeness of information was also obtained through observation by conducting interviews with one of the Banten Provincial Statistics through BPS(Badan Pusat Statistik) social media as shown in table 1.

| District/City         | Life Expectancy According to District/City |
|-----------------------|--------------------------------------------|
|                       | 2013 | 2014 | 2015 | 2016 | 2017 |
| Pandeglang district   | 62.83| 62.91| 63.51| 63.77| 64.04|
| Lebak district        | 65.83| 65.88| 66.28| 66.43| 66.59|
| Tangerang district    | 68.96| 68.98| 69.28| 69.37| 69.47|
| Serang district       | 63.03| 63.09| 63.59| 63.81| 64.02|
| Tangerang city        | 71.09| 71.09| 71.29| 71.34| 71.38|
| Cilegon city          | 65.84| 65.85| 66.15| 66.24| 66.32|
3.2 Analysis Techniques

Analysis technique presented in this discussion is an overall description of the processes that are based on the formulation of the problem as shown in figure 1.

**Figure 1. Stages of Analysis Techniques**

In Figure 2, it is explained that researchers need all category indicator data for each region in 2013 to 2015. Then the data will be processed using the K-Means algorithm. The author classifies the data using the K-Means algorithm, then the results show information on regional groups according to the criteria.

**Figure 2. Stages of K-Means Algorithm**

| Location            | 2013   | 2014   | 2015   | 2016   | 2017   |
|---------------------|--------|--------|--------|--------|--------|
| Serang city         | 67.23  | 67.23  | 67.33  | 67.36  | 67.38  |
| South Tangerang city | 72.10  | 72.11  | 72.12  | 72.14  | 72.16  |
| Kep. Seribu district | 67.16  | 67.22  | 67.72  | 67.88  | 68.04  |
| South Jakarta city  | 73.80  | 73.81  | 73.81  | 73.83  | 73.84  |
| East Jakarta city   | 73.98  | 74.00  | 74.10  | 74.14  | 74.18  |
| Center Jakarta city | 73.58  | 73.60  | 73.70  | 73.76  | 73.83  |
| West Jakarta city   | 73.22  | 73.22  | 73.32  | 73.34  | 73.37  |
| North Jakarta city  | 72.80  | 72.81  | 72.91  | 72.95  | 72.99  |
3.3 K-Means Calculation Process

The following is an example of the grouping calculation process on life expectancy, the stages will be directly transferred to the last iteration, \( r \), the data used can be seen in Table 1.

| Table 2. Centroid Iteration Results (Last) |
|-------------------------------------------|
| **Year / Cluster** | 2013 | 2014 | 2015 | 2016 | 2017 |
|---------------------|------|------|------|------|------|
| C1                  | 62.92888 | 63.00388 | 63.55388 | 63.79 | 64.03 |
| C2                  | 67.00332 | 67.03132 | 67.35132 | 67.456 | 67.56 |
| C3                  | 72.93832 | 72.9476 | 73.03564 | 73.07143 | 73.10714 |

Because the position of the centroid no longer changes, the iteration process is complete. Table 2 is the value of the final centroid, which will be useful to be used as a reference for the level of group values. From the results obtained, C1 becomes the group that has the lowest value, C2 becomes the group that has a moderate value, and C3 becomes the group that has the highest value.

4. Experimental Results

4.1 Grouping Results

From the results of calculations using the K-Means algorithm. The following are tables and graphs of the results of grouping a region, the tables and graphs below consist of indicators of life expectancy, school length expectations, average length of schooling, and per capita expenditure:

| Table 3. Central Point of Life Expectancy |
|-----------------------------------------|
| **Year / Cluster** | 2013 | 2014 | 2015 | 2016 | 2017 | Group Status |
|---------------------|------|------|------|------|------|--------------|
| C1                  | 62.92888 | 63.00388 | 63.55388 | 63.79 | 64.03 | Low group |
| C2                  | 67.00332 | 67.03132 | 67.35132 | 67.456 | 67.56 | Medium group |
| C3                  | 72.93832 | 72.9476 | 73.03564 | 73.07143 | 73.10714 | High group |

The table 3 is the result of calculating the central point of life expectancy. The table will be a reference for the level of group value. From the results obtained, C1 becomes the group that has the lowest value, C2 becomes the group that has a medium value, and C3 becomes the group that has the highest value.

| Table 4. Regional Grouping Results Based on Life Expectancy Figures |
|---------------------------------------------------------------|
| **District / City** | Life Expectancy According to District/City | Group |
|---------------------|-------------------------------------------|-------|
|                     | 2013 | 2014 | 2015 | 2016 | 2017 | With |
| Pandeglang district | 62.83 | 62.91 | 63.51 | 63.77 | 64.04 | Low value |
| Lebak district      | 65.83 | 65.88 | 66.28 | 66.43 | 66.59 | Medium value |
| Tangerang district  | 68.96 | 68.98 | 69.28 | 69.37 | 69.47 | Medium value |
| Serang district     | 63.03 | 63.09 | 63.59 | 63.81 | 64.02 | Low value |
| Tangerang city      | 71.09 | 71.09 | 71.29 | 71.34 | 71.38 | High value |
Based on the results on Table 4, the Pandeglang district and the attack district are among the regions with the smallest value, meaning that the area has a minimum life expectancy value between Banten Province and DKI Jakarta Province. Lebak Regency, Tangerang, Seribu Islands, Cilegon City and Serang City are among the regions with moderate values. The cities of Tangerang, South Tangerang, South Jakarta, East Jakarta, Central Jakarta, West Jakarta and North Jakarta are among the regions with the highest scores.

**Table 5. Central Points for School Length Expectancy**

| Year / Cluster | 2013  | 2014        | 2015       | 2016        | 2017       | Group Status          |
|---------------|-------|-------------|------------|-------------|-----------------------|
| C1            | 12.76329 | 13.078248  | 13.16212   | 13.35857    | 13.438571            | High group             |
| C2            | 11.81987 | 12.11467    | 12.12827   | 12.14       | 12.15                | Medium group           |
| C3            | 11.68337 | 12.03026    | 12.2326    | 12.412      | 12.56                | Low group              |

Table 5 is the result of calculating the central point for school length expectancy. The table will be a reference for the level of group value. From the results obtained, C1 becomes the group that has the lowest value, C2 becomes the group that has a medium value, and C3 becomes the group that has the highest value.

**Table 6. Regional Grouping Results Table Based on School Long Expectations**

| District / City      | Life Expectancy According to District/City | Group     |
|----------------------|-------------------------------------------|-----------|
|                      | 2013 | 2014 | 2015 | 2016 | 2017 |                      |
| Pandeglang district  | 12.86 | 13.38 | 13.39 | 13.40 | 13.41 | High value          |
| Lebak district       | 11.55 | 11.88 | 11.90 | 11.91 | 11.92 | Medium value        |
| Tangerang district   | 11.44 | 11.65 | 11.89 | 12.11 | 12.51 | Low value           |
| Serang district      | 12.09 | 12.35 | 12.36 | 12.37 | 12.38 | Medium value        |
| Serang city          | 12.60 | 12.86 | 12.90 | 13.41 | 13.44 | High value          |
| Cilegon city         | 12.67 | 13.07 | 13.10 | 13.11 | 13.12 | High value          |
| Serang city          | 11.92 | 12.34 | 12.36 | 12.63 | 12.64 | Low value           |
| South Tangerang city | 13.24 | 13.58 | 13.61 | 14.08 | 14.39 | High value          |
| Kep. Seribu district | 11.21 | 11.89 | 11.90 | 12.10 | 12.40 | Low value           |
| South Jakarta city   | 12.96 | 13.09 | 13.09 | 13.22 | 13.27 | High value          |
| East Jakarta city    | 12.63 | 13.06 | 13.08 | 13.20 | 13.26 | High value          |
| Center Jakarta city  | 12.39 | 12.51 | 12.96 | 13.09 | 13.18 | High value          |
| West Jakarta city    | 12.01 | 12.39 | 12.60 | 12.69 | 12.70 | Low value           |
| North Jakarta city   | 11.84 | 11.89 | 12.42 | 12.53 | 12.55 | Low value           |
Based on the results on table 6, Tangerang Regency, Tangerang, Thousand Islands, West Jakarta City, and North Jakarta City are among the smallest value groups, meaning that the area has minimal long-term school expectations between Banten Province and DKI Jakarta Province. Lebak Regency, and attack includes a group of regions with moderate values. Pandeglang Regency, Tangerang City, Cilegon City and Serang City, South Tangerang City, South Jakarta City, East Jakarta City, Central Jakarta City are among the highest value regions.

Table 7. Central Point of Averange of School Length

| Year / Cluster | 2013    | 2014    | 2015    | 2016    | 2017    | Group Status     |
|----------------|---------|---------|---------|---------|---------|------------------|
| C2             | 6.299396| 6.326086| 6.455228| 6.596667| 6.666667| Low group        |
| C1             | 8.244693| 8.270709| 8.28682 | 8.356667| 8.366667| Medium group     |
| C3             | 10.46993| 10.55574| 10.63412| 10.76   | 10.85125| High group       |

Table 7 is the result of calculating the central point of average of school length. The table will be a reference for the level of group value. From the results obtained, C2 becomes the group that has the lowest value, C1 becomes the group that has a medium value, and C3 becomes the group that has the highest value.

Table 8. Regional Grouping Results Table Based on Average School Length

| District / City          | Life Expectancy According to District/City | Group             |
|--------------------------|-------------------------------------------|-------------------|
|                          | 2013 | 2014 | 2015 | 2016 | 2017 | With              |
| Pandeglang district      | 6.44 | 6.45 | 6.60 | 6.62 | 6.63 | Low value         |
| Lebak district           | 5.81 | 5.84 | 5.86 | 6.19 | 6.20 | Low value         |
| Tangerang district       | 8.18 | 8.20 | 8.22 | 8.23 | 8.24 | Medium value      |
| Serang district          | 6.65 | 6.69 | 6.90 | 6.98 | 7.17 | Low value         |
| Tangerang city           | 9.82 | 10.20| 10.20| 10.28| 10.29| High value        |
| Cilegon city             | 9.60 | 9.66 | 9.67 | 9.68 | 9.69 | High value        |
| Serang city              | 8.56 | 8.58 | 8.59 | 8.60 | 8.61 | Medium value      |
| South Tangerang city     | 11.48| 11.56| 11.57| 11.58| 11.77| High value        |
| Kep. Seribu district     | 7.99 | 8.03 | 8.04 | 8.24 | 8.25 | Medium value      |
| South Jakarta city       | 10.95| 10.97| 11.23| 11.42| 11.47| High value        |
| East Jakarta city        | 11.18| 11.21| 11.32| 11.52| 11.60| High value        |
| Center Jakarta city      | 10.85| 10.87| 10.88| 11.01| 11.02| High value        |
| West Jakarta city        | 10.04| 10.13| 10.15| 10.36| 10.37| High value        |
| North Jakarta city       | 9.85 | 9.85 | 10.05| 10.23| 10.60| High value        |

Based on the result on table 8, the regency of Pandeglang, Lebak, Serang, belongs to the group with the smallest value, meaning that the area has a minimum value of old school expectations between Banten Province and DKI Jakarta Province. Regency of Tangerang Regency, Serang City, Thousand Islands, including groups with moderate values. Tangerang City District, Cilegon, South Tangerang, South Jakarta, Jakarta Timur, Jakarta Pusat, West Jakarta, East Jakarta, including the highest value groups in the region.
Table 9. Central Point of Expenditure Per capita

| Year / Cluster | 2013     | 2014     | 2015     | 2016     | 2017     | Group Status   |
|---------------|----------|----------|----------|----------|----------|----------------|
| C1            | 62.92888 | 63.00388 | 63.55388 | 63.79    | 64.03    | Low group      |
| C2            | 67.00332 | 67.03132 | 67.35132 | 67.456   | 67.56    | Medium group   |
| C3            | 72.93832 | 72.9476  | 73.03564 | 73.07143 | 73.10714 | High group     |

Table 9 is the result of calculating the central point of expenditure per capita. The table will be a reference for the level of group value. From the results obtained, C1 becomes the group that has the lowest value, C2 becomes the group that has a medium value, and C3 becomes the group that has the highest value.

Table 10. Table of Regional Grouping Results Based on Per capita Expenditures

| District / City          | Life Expectancy According to District/City | Group          | With          |
|--------------------------|--------------------------------------------|----------------|---------------|
|                          | 2013 | 2014 | 2015     | 2016     | 2017     |                |
| Pandeglang district      | 7486 | 7589 | 7730     | 8138    | 8358     | Low value      |
| Lebak district           | 7918 | 7977 | 8111     | 8308    | 8372     | Low value      |
| Tangerang district       | 11648| 11666| 11727    | 11865   | 11914    | Medium value   |
| Serang district          | 9831 | 9886 | 10004    | 10317   | 10466    | Low value      |
| Tangerang city           | 13531| 13671| 13766    | 13911   | 14104    | Medium value   |
| Cilegon city             | 11920| 12057| 12127    | 12326   | 12562    | Medium value   |
| Serang city              | 11950| 12091| 12289    | 12660   | 12914    | Medium value   |
| South Tangerang city     | 14207| 14361| 14588    | 14972   | 15291    | Medium value   |
| Kep. Seribu district     | 11253| 11316| 11433    | 11608   | 11833    | Medium value   |
| South Jakarta city       | 22067| 22208| 22425    | 22932   | 23098    | High value     |
| East Jakarta city        | 16181| 16248| 16455    | 16733   | 17007    | High value     |
| Center Jakarta city      | 15820| 15922| 16143    | 16493   | 16719    | High value     |
| West Jakarta city        | 18794| 18897| 19006    | 19501   | 19695    | High value     |
| North Jakarta city       | 16851| 16959| 17205    | 17418   | 17763    | High value     |

Table 10 is the results of the grouping, explained that Pandeglang and Serang regencies are regions that are always in the three lowest groups, namely indicators of life expectancy with a central point of 62.9288 in 2013 up to 64.03 in 2017, the average length of school with a central point 6.2993 of 2013 up to 6.666 in 2017, per capita expenditure with a center point of 8411.72 in 2013 to 9605.33 in 2017.

4.2 Test Results
From the results of the trial testing the accuracy of the K-Means value, the researcher must prepare a label that is available at the centroid point, namely, '0' and '1'. The centroid point is the result of grouping objects based on the closest distance. The following are the results of accuracy testing on the value of life expectancy, long school expectations, average length of schooling, and expenditure per capita using purity:

Table 11. Central Point of Life Expectancy

| Central Point | C1 | C2 | C3 |
|---------------|----|----|----|
|               | 1  | 0  | 0  |
|               | 0  | 1  | 0  |
The test was carried out using equation (4), the closer the value of ‘0’, the worse the results, the closer to ‘1’, the better the cluster value.

Table 12. Table of Testing Method Results on Life Expectancy Figures

| Cluster | Label | Total (Label) | Total of Data | Max | Purity Result |
|---------|-------|---------------|---------------|-----|---------------|
| C1      | 0     | 12            | 14            | 12  | 0.857142857   |
|         | 1     | 2             |               |     |               |
| C2      | 0     | 9             | 9             | 9   | 0.642857143   |
|         | 1     | 5             |               |     |               |
| C3      | 0     | 7             | 7             | 7   | 0.5           |
|         | 1     | 7             |               |     |               |

Table 13. Points Length of Study Index

| Central Point | C1 | C2 | C3 |
|---------------|----|----|----|
| C1            | 1  | 0  | 0  |
| C2            | 0  | 1  | 0  |
| C3            | 0  | 0  | 1  |
| C4            | 1  | 0  | 0  |
| C5            | 0  | 0  | 1  |
| C6            | 1  | 0  | 0  |
| C7            | 0  | 0  | 1  |
| C8            | 1  | 0  | 0  |
| C9            | 0  | 0  | 1  |
| C10           | 0  | 0  | 1  |

Table 14. Table of Results of Testing Methods at School Long Expectations

| Cluster | Label | Sum of Label | Total of Data | Max | Purity Result |
|---------|-------|--------------|---------------|-----|---------------|
| C1      | 0     | 7            | 7             | 7   | 0.5           |
|         | 1     | 7            |               |     |               |
| C2      | 0     | 12           | 12            | 12  | 0.857142857   |
|         | 1     | 2            |               |     |               |
| C3      | 0     | 9            | 9             | 9   | 0.642857143   |
|         | 1     | 5            |               |     |               |
Table 15. Average Center for Length of Study Index

| Central Point | C1 | C2 | C3 |
|---------------|----|----|----|
| 0             | 1  | 0  | 0  |
| 0             | 1  | 0  | 0  |
| 1             | 0  | 0  | 0  |
| 0             | 1  | 0  | 0  |
| 0             | 0  | 1  | 0  |
| 0             | 0  | 0  | 1  |
| 1             | 0  | 0  | 0  |
| 0             | 0  | 1  | 0  |
| 1             | 0  | 0  | 0  |
| 0             | 0  | 0  | 1  |
| 1             | 0  | 0  | 0  |
| 0             | 0  | 1  | 0  |
| 0             | 0  | 0  | 1  |
| 0             | 0  | 1  | 0  |
| 0             | 0  | 0  | 1  |

Table 16. Testing Methods Results on Average School Length

| Cluster | Label | Sum of label | Total of data | Max Purity Result |
|---------|-------|--------------|---------------|-------------------|
| C1      | 0     | 11           | 11            | 0.785714286       |
|         | 1     | 3            |               |                   |
| C2      | 0     | 11           | 14            | 0.785714286       |
|         | 1     | 3            |               |                   |
| C3      | 0     | 6            | 8             | 0.571428571       |
|         | 1     | 8            |               |                   |

Table 17. Per capita Expenditure Center Tab

| Central Point | C1 | C2 | C3 |
|---------------|----|----|----|
| 1             | 0  | 0  | 0  |
| 1             | 0  | 0  | 0  |
| 0             | 1  | 0  | 0  |
| 1             | 0  | 0  | 0  |
| 0             | 1  | 0  | 0  |
| 0             | 1  | 0  | 0  |
| 0             | 1  | 0  | 0  |
| 0             | 0  | 1  | 0  |
| 0             | 0  | 1  | 0  |
| 0             | 0  | 1  | 0  |
| 0             | 0  | 1  | 0  |

Table 18. Results of Testing Methods on Perkapita Expenditures

| Cluster | Label | Sum of Label | Total of data | Max Purity Result |
|---------|-------|--------------|---------------|-------------------|
| C1      | 0     | 11           | 14            | 0.785714286       |
|         | 1     | 3            |               |                   |
| C2      | 0     | 8            | 8             | 0.571428571       |
|         | 1     | 6            |               |                   |
| C3      | 0     | 9            | 9             | 0.642857143       |
|         | 1     | 5            |               |                   |

It can be concluded that datasets with data attributes. The K-Means algorithm uses the mean as the center of the centroid, also Euclidean as a function of distance to calculate the distance of proximity between objects with the centroid.
5. Conclusion
Based on the results of research carried out from the initial stage until the testing process of clustering application, the level of basic human category indicators can be taken several conclusions, the K-Means algorithm is used to classify an area based on indicators of basic human categories (life expectancy, school length expectations, average school length, expenditure per capita) which refers to the value of the center point (centroid). The group results from this study are, areas with low scores, areas with moderate values, and areas with high scores. Pandeglang and Serang Regencies are areas that are always in the three lowest groups such as life expectancy indicators with a central point of 62.9288 in 2013 to 64.03 in 2017, the average length of school with a central point of 6.2993 in 2013 up to 6.666 in 2017, and per capita expenditure with a center point of 8411.72 in 2013 to 9605.33 in 2017. South Jakarta City, East Jakarta, and Central Jakarta are areas that are always in the highest group in all indicators. The center point of life expectancy is 72.938 in 2013 to 73.107 in 2017, the center point of school long-term expectations is 12.763 in 2013 to 13.438 in 2017, the center point of the average school length is 10.469 in 2013 up to 10.851 in 2017, the center point of expenditure per capita is 17942.57 of 2013 to 18856.4 in 2017.

References
[1] Badan Pusat Statistik Provinsi Jakarta, *IPM DKI Jakarta Tahun 2017 Masuk Level “Sangat Tinggi”*, Badan Pusat Statistik, Jakarta, 2017.
[2] Badan Pusat Statistik Provinsi Banten, *Indeks Pembangunan Manusia Provinsi Banten 2016*, Badan Pusat Statistik, 2016, 2016.
[3] Badan Pusat Statistik, *Indeks Pembangunan Manusia 2016*, Jakarta: Badan Pusat Statistik, 2016.
[4] V. K. Singh, N. Tiwari och S. Garg, *Document Clustering using K-means, Heuristic K-means and Fuzzy C-means*, IEEE, pp. 297-301, 2011.
[5] Y. A. Priambodo och S. Y. J. Prasetyo, *Pemetaan Penyebaran Guru di Provinsi Banten dengan Menggunakan Metode Spatial Clustering K-Means (Studi kasus : Wilayah Provinsi Banten)*, Indonesian Journal of Computing and Modeling, pp. 18-27, 2018.
[6] S. Natural Language Processing Group, *Evaluation of Clustering*, 07 04 2009. [Online]. Available: https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-clustering-1.html.