Evaluation of scoring method in urban-rural classification in Indonesia

Herlina¹, R P Manik¹, B Sartono¹* and L Suharni²

¹Statistics Department, IPB University, Bogor, 16680, Indonesia
²BPS-Statistics Indonesia, Jakarta, 10710, Indonesia

*E-mail: bagusco@ipb.ac.id

Abstract. BPS-Statistics Indonesia is responsible for collecting valid official data in Indonesia, which mainly depends on the sampling plan they used. The village classification into urban or rural group plays a significant role in the construction of the plan. So far, the classification has been based on a scorecard developed in 2010 consisting of 12 variables. With a certain threshold, each variable has several criteria that give a score. The total sum of the scores is then used to define the class of urban and rural. This study worked explorative to evaluate whether the scorecard is still proper to be implemented. Exploratory data analysis used in this study to analyzing data structures and patterns of the data distribution. Both of them will uncover unexpected things from the data. The result shows that variables i.e. population density, the percentage of agricultural households, the existence of kindergarten, the existence of cinema, the existence of hotel/pub/beauty shop, the percentage of household with cable phone and the percentage of household with electricity require more sensitive criteria and scorecard.

1. Introduction
BPS-Statistics Indonesia is the official statistical institution in Indonesia that is responsible for collecting valid official data in order to measure national development. In general, the data collected is sourced from surveys that require a sampling design. One of the concerns in sampling design is the representation of urban and rural areas. Villages, as the smallest administrative unit in Indonesia, are classified as an urban or rural area.

Unlike the term cities and villages, the term urban and rural areas refer to the characteristics that are attached to an area. The characteristic of an urban and rural area is different. UN defines that rural related to an area with characteristic lower population density, lack of basic infrastructure (road, access to electricity, water, telephone, school, etc.) [1]. United Nation also states that not possible to adopt uniform criteria to distinguish urban from rural given the variety situation in the countries of the world. [1] Research needed to determine appropriate criteria to characterize the urban and rural area.

Currently, urban-rural classification by BPS uses the 2010 scoring method which uses 12 variables. Scoring to classify villages into urban-rural areas based on the variable i.e. the population density, the percentage of agricultural households, and the existence/access to reach some urban facilities that called as access to the urban facility [2].

The scoring method can be described as the process of scoring the criteria of each classification variable. Scoring results are summed to the total score and compared to the threshold of 10. If the total score is 10 or more then the village is classified as an urban village, and vice versa [2]. The threshold of 10 is a value determined based on the previous internal research by BPS-Statistics Indonesia.
Table 1. Guidelines to urban-rural classification with criteria and scores in 2010

| Criteria          | Score | Score | Urban Facility | Criteria | Score |
|-------------------|-------|-------|----------------|----------|-------|
| Population Density|       |       |                |          |       |
| <500              | 1     | >70.00| a. Kindergarten|          |       |
| 500-1,249         | 2     | 50.00-69.90| b. Junior High School| - Have or 2.5 km | 1 |
| 1,250-2,499       | 3     | 30.00-49.99| c. Senior High School| - > 2.5 km | 0 |
| 2,500-3,999       | 4     | 20.00-29.99| d. Traditional Market| - Have or ≤ 2 km | 1 |
| 4,000-5,999       | 5     | 15.00-19.99| e. Mall/Shopping Complex| - > 2 km | 0 |
| 6,000-7,499       | 6     | 10.00-14.99| f. Cinema| - Have or ≤ 5 km | 1 |
| 7,500-8,499       | 7     | 5.00-9.99| g. Hospital| - > 5 km | 0 |
| > 8,500           | 8     | < 5.00| h. Hotel/Pub/Beauty shop| - Have | 1 |
|                   |       |       |                | - Not Have | 0 |
|                   |       |       | i. Percentage of household with cable phone| - ≥ 8.00 | 1 |
|                   |       |       |                | - < 8.00 | 0 |
|                   |       |       | j. Percentage of household with electricity| - ≥ 90.00 | 1 |
|                   |       |       |                | - < 90.00 | 0 |

*BPS (2010)*

The development of information and transportation technology makes it easier for villages to be actively involved in economic activities, public services and other activities that are generally located in the city. For example, the presence of a 24-hour clinic makes the villagers not have to go to the city to get adequate health services. Or, the management of village tourism independently makes villages have new economic resources besides the agricultural sector. Finally, there is a phenomenon in Indonesia that the development of cities characteristics increases in rural areas which have implications for increasing urban characterization in rural areas [3]. The trend of globalization and urbanization affects villages to take on urban elements. A rural area that may currently be affected by the urbanization process, make it possible to become urban in the future.

Research relating to the evaluation of rural-urban classifications in Indonesia is still limited. [3] did an evaluation of the village development in Jawa Barat by remapping the typology of the village. This research strengthens the opinion that villages develop from rural to be urban.

Urban and rural developments can be seen from data changes from year to year. One tool that can be used is the Exploratory Data Analysis (EDA) to see urban and rural developments. EDA is an approach in analyzing data sets to summarize the main characteristics of data, including through visual methods. EDA is the first step after data collection and pre-processing, where the data is simply visualized, plotted, manipulated, without any assumptions so that it is useful in guiding data quality and building models [3]. Some of the objectives of data exploration analysis include: maximizing data analysis, finding data structures, detecting important variables, including detecting outliers and anomalies and visualizing potential relationships between response and predictor variables [3].

The method of analyzing data exploration is cross-classified between graph usage (graphical or non-graphical) and the type of variable (univariate or multivariate). In this research, EDA is used to observe
total score data and inter-year predictor variables, then take a descriptive measure as a summary value of existing observation values. The Kolmogorov Smirnov result test can be done to support EDA output.

This study worked explorative to evaluate whether the scorecard is still proper to be implemented. In this study, Exploratory Data Analysis is used to analyzing data structures and patterns of the data distribution. This tool will uncover things from the data.

Materials
The variables in the 2010 scoring method are used in this paper (Table 2) [2]. The dataset of this work was obtained from the raw data on *Pendataan Potensi Desa* (PODES) 2008, 2011, and 2014. The observation unit in this study is all villages in Indonesia that have existed since 2008 until 2014.

| No. | Variable                          | Information                                      |
|-----|-----------------------------------|--------------------------------------------------|
| 1   | Population density (X1)           | Variables that conform to the 2010 scoring method guidelines. |
| 2   | Percentage of agricultural household (X2) |
| 3   | The existence of Kindergarten (X3) |
| 4   | The existence of Junior High School (X4) |
| 5   | The existence of Senior High School SMA (X5) |
| 6   | Traditional market (X6)           |
| 7   | Mall/shopping complex (X7)        |
| 8   | Cinema (X8)                       |
| 9   | Hospital (X9)                     |
| 10  | Hotel/Pub/Beauty shop (X10)       |
| 11  | Percentage of household with cable phone (X11) |
| 12  | Percentage of household with electricity (X12) |

2. Methods
Pre-processing by cleaning stages are carried out to get the observation unit. Cleaning data includes filtering the number of villages that exist in all period of a dataset. This applied with matching village by the id Desa, the unique code of villages in the Villages Framework. The next process of data cleaning is to select question items according to the variables used in determining urban or rural status based on the 2010 scoring method, as stated in Table 1.

Three important things that must be considered in this study are the predictor variable, the criteria of the predictor variable and the data distribution pattern in each predictor. Predictor variables are independent variables to distinguish villages according to the urban or rural status. The score of each criterion is the weight of each criterion in each predictor. Data distribution patterns will provide information about the comparison of data between years of observation, whether it has similar or different data distribution patterns. These things become a sensitive factor because of villages development from year to year.

This paper uses the univariate method, both with graphical and non-graphical. The graph that will be used is the line plot of the data density function. The non-graphical method used is the contingency table and the characteristics of quantitative data such as the measure of central tendency and shape of the distribution. This measure is needed as a summary value of existing observation values [4]. Some of the measures used in this paper are mean, median, quintile, the minimum value and the maximum value (summarized as five statistical series).
The Exploratory Data Analysis (EDA) in this paper as follows:

- Applying the 2010 scoring method to 2008, 2011 and 2014 datasets. The result is summarized in the transitions matrix of the urban-rural status of villages between 2008-2011 and 2008-2014 (Table 3).

Table 3. Transition matrix designs of the urban-rural status of the villages by applying 2010 scoring method (2008 and 2014 as an example)

|        | Rural | Urban | Total |
|--------|-------|-------|-------|
| 2008 Urban | c     | d     |       |
| 2014 Rural | a     | b     |       |
| Total   |       |       |       |

- Showing measure of central tendency, i.e. the mean score for each variable. These are applied to 12 variables. A variable whose small variation may not differentiate the status among the village, therefore if we found variable like this so we propose to modify or replace the variables or the criteria.

- Showing data distribution of the variable.

   By paying attention to the pattern of data distribution in each predictor for each data period, will show similar or different data distribution patterns. If different indicates the classification is less relevant to the current situation so it is necessary to modify or replace variables and criteria. For this part, The Kolmogorov Smirnov result test can be done to support EDA output.

   Kolmogorov Smirnov Test (KS-test) is a statistical measurement about the distance between the sample distribution function empirically and the cumulative distribution function of the reference distribution or measuring the distance between the empirical distribution function of the two samples. K-S test measures the proximity of the distance $F_{1,n}(x)$ with $F_{2,m}(x)$ to test whether two underlying one-dimensional probability distributions differ:

$$D_{n,m} = \sup_{-\infty < x < \infty} \left| F_{1,n}(x) - F_{2,m}(x) \right|$$

where:

- $D_{n,m}$ : the furthest vertical distance between $F_{1,n}(x)$ and $F_{2,m}(x)$
- $D^*(\alpha)$ : critical values from the Kolmogorov Smirnov test table.

3. Results and Discussions

3.1. Development of the urban-rural status

The datasets in three periods have different numbers of villages, therefore we applied matching villages by the id desa. The villages that use in this study is the villages that exist in all periods of a dataset. Matching village by id Desa results in as many as 66,902 villages that exist in three period’s dataset.

We applied the scoring method of Table 1 to the 66,902 villages so that every village was categories into urban or rural status. The cross-tabulation of number villages by urban-rural status in 2008 and
2011 dataset was represented as transition matrix in Table 4. Then in Table 5 is the cross-tabulation of number villages by urban-rural status in 2008 and 2014 dataset.

### Table 4. Transition matrix of the urban-rural status of the villages in 2008 and 2011

|        | 2011 | Total |
|--------|------|-------|
|        | Rural| Urban |
| 2008   | 49 016| 4 499 | 53 515 |
|        | 1 771| 11 616| 13 387 |
| Total  | 50 787| 16 115| 66 902 |

### Table 5. Transition matrix of the urban-rural status of the villages in 2008 and 2014

|        | 2014 | Total |
|--------|------|-------|
|        | Rural| Urban |
| 2008   | 47 412| 6 103 | 53 515 |
|        | 1 266| 12 121| 13 387 |
| Total  | 48 678| 18 224| 66 902 |

Based on tables and 4 and 5 it is known that the number of villages with urban status has increased from 2008, 2011 and 2014, and vice versa with the number of villages with rural status. In 2008, the number of villages with urban status was 13,387 villages, 2011 increased to 16,115 and 2014 as many as 18,224 urban villages.

However, it is known that there are still a number of villages that have changed their status from urban villages to rural villages. In 2011 there were 1,771 rural villages which in 2008 were urban. This also happened in 2014, where there were 1,266 rural villages even though previously had urban status.

This shows that the current scoring method is in line with the initial assumption that the increasing number of villages with rural to urban status, and there is a small change in the number of urban villages to rural villages, even the value is close to zero. However, the hope is that there will be fewer villages that change status from urban to rural status. Thus it can be stated that it is necessary to examine existing assessment methods.

3.2. The mean score of the variable

Table 6 present the mean score of all variable used to classification urban-rural status with the scoring method. We use the table to identify which variable needs to be considered to modification. This could be done with several circumstances. First, it possible that the variable to be extremely homogeneous. Second conditions that the variable may not be consistent in time.

Let’s look at X8 (the existence of cinema). Each village was score zero (0) if there is not any cinema in the area and score 1 when there exists at least one cinema or if there is no cinema but the distance to the nearest cinema can access up to 5 kilometers. The mean score of X8 is 0.0357 (in 2008) which means only 3.57 percent of villages having the cinema or almost all villages has no cinema. And it happens to in 2011 and 2014 that the mean score as 0.0336 and 0.0335. It indicates the existence of cinema could not be useful to identify urban or rural status. A pattern similar to X8 can be seen in X10 (the existence of a hotel/pub/beauty shop) and X11 (the Percentage of household with cable phone).

The mean score of X11 in 2008, 2011 and 2014 showed a decrease wherein 2008 have 12.96 percent of villages that households used cable phone, but in 2014 there was smaller the number of villages, that was only 4.90 percent of the villages whose households used the cable phone. It is shown that the percentage of households with cable phone is less identifying urban or rural status. The increasing use
of smartphones is considered one of the causes of this variable to be less relevant in determining urban or rural status.

Table 6. The mean score of the variable in 2008, 2011 and 2014

| Variable Index | Variable                                      | Score Range | 2008     | 2011     | 2014     |
|----------------|-----------------------------------------------|-------------|----------|----------|----------|
| X1             | Population density                           | 1-8         | 1.9429   | 1.8831   | 1.9610   |
| X2             | Percentage of agricultural household         | 1-8         | 1.8950   | 2.3984   | 2.4719   |
| X3             | The existence of kindergarten                 | 0;1         | 0.7233   | 0.7669   | 0.8077   |
| X4             | The existence of the junior high school       | 0;1         | 0.6287   | 0.6719   | 0.7086   |
| X5             | The existence of the senior high school       | 0;1         | 0.3846   | 0.3006   | 0.4648   |
| X6             | The existence of the traditional market       | 0;1         | 0.3195   | 0.3408   | 0.3544   |
| X7             | The existence of the mall/shopping complex   | 0;1         | 0.2266   | 0.2467   | 0.2554   |
| X8             | The existence of cinema                      | 0;1         | 0.0357   | 0.0336   | 0.0335   |
| X9             | The existence of the hospital                 | 0;1         | 0.1612   | 0.1699   | 0.1798   |
| X10            | The existence of a hotel/pub/beauty shop      | 0;1         | 0.0757   | 0.0815   | 0.0906   |
| X11            | Percentage of household with cable phone     | 0;1         | 0.1296   | 0.0852   | 0.0490   |
| X12            | Percentage of household with electricity      | 0;1         | 0.5316   | 0.6917   | 0.8088   |

The variables of X3 and X12 with the mean score that closes to 1. For X3 wherein 2014, the mean score is 0.8077 which means 80.77 percent of villages already have kindergartens or if there are no kindergartens but the distance to the nearest kindergartens can access up to 2.5 kilometers. It means that almost of villages have kindergartens.

Variable X12 a value close to 1 means that more villages with 90 percent of households in the village have electricity. The variable X12 shows the movement of the mean score that increases and closes to 1. The mean score in 2008 was 0.5316, then in 2011 was 0.6917 and in 2014 was 0.8088. This shows that almost villages with 90 percent of households have used electricity, thus this variable is not useful in determining urban or rural status.

The variables X3 and X12 have a score range from 1 to 8. The X1 variable in all periods had a mean score of 2, meaning that almost all villages have a population density of around 500 to 2149 people per square kilometer. In the X2, the mean score rises from 1.89 in 2008, 2.39 in 2011 and to 2.47 in 2014. This shows that from year to year there was a decrease in the percentage of agricultural households in the village.

The interesting to be observed from the mean score on the variables X1 and X2. The mean score ranges from criterion 2, even though there are 8 criteria. This raises the idea that the criteria for these variables are not appropriate in identifying urban or rural status.

3.3. Distribution of the variable

Plot R1 - R12 (Figure 1) is a line plot of the distribution data of variable X1 until X12 (except X10). R1 shows a line plot of population density. R2 shows a line plot of the percentage of agricultural households. Then, R3-R9 represents a line plot of the distance to access public facilities (X3-X9). Furthermore, R11 and R12 represent a line plot of the density function of X11 and X12. The blue line shows the 2008 dataset and the red line for the 2014 dataset.

The images show different patterns of change between 2008 and 2014. Changes in the data distribution pattern indicate that outcomes from the national developments between 2008 and 2014. From Figure 1, R1, R2, R11, and R12 have different patterns between 2008 and 2014. As with R3, R4, R5, R6, R7, R8 and R9 which have almost the same pattern.
Figure 1. The comparison of data distribution between 2008 and 2014 datasets

- a. Population Density (R1)
- b. Percentage of Agricultural Household (R2)
- c. The closest distance to kindergarten (R3)
- d. The closest distance to junior high school (R4)
- e. The closest distance to senior high school (R5)
- f. The closest distance to the traditional market (R6)
- g. The closest distance to the mall/shopping complex (R7)
- h. The closest distance to Cinema (R8)
- i. The closest distance to Hospital (R9)
- j. The closest distance to Hospital (R9)
- k. Percentage of household with cable phone (R11)
- l. Percentage of household electricity (R12)
Table 7 shown analysis non-graphical, that present the summary statistics of a dataset in 2008 and 2014. This table presents minimum value, quintile 1, median, mean, quintile 3 and maximum value. Median is the value of the data that divides into half as many of the sets of data that have been sorted. The mean and median have the same value, it indicates normal data distribution. If the mean is greater than the median, indicates the right skewness data distribution, and vice versa.

The quartile Q1 – Q4, divides a sorted observation data into four parts, each of which includes a quarter of all observational data. The distance between Q3 and Q1 is called the interquartile range (IQR). IQR is covering 50 percent of the data around the median. Long IQR shows that data tend to be spread or have a high distribution size. In contrast, short IQR shows that data observation has a low spread data distribution.

| Index | Variable | Year | Min  | Q1   | Median | Mean   | Q3   | Max  |
|-------|----------|------|------|------|--------|--------|------|------|
| R1    | Population density | 2008 | 0.00 | 77.90 | 359.40 | 1 334.10 | 1 157.60 | 66 575.00 |
|       |          | 2014 | 0.00 | 78.97 | 347.25 | 1 484.03 | 1 166.55 | 62 642.42 |
| R2    | Percentage of agricultural household | 2008 | 0.00 | 60.00 | 80.00 | 71.58 | 92.00 | 99.00 |
|       |          | 2014 | 0.00 | 37.93 | 71.85 | 63.14 | 91.98 | 100.00 |
| R3    | The closest distance to kindergarten | 2008 | 0.00 | 0.00 | 0.00 | 5.73 | 3.00 | 99.80 |
|       |          | 2014 | 0.00 | 0.00 | 0.00 | 4.42 | 1.00 | 99.90 |
| R4    | The closest distance to junior high school | 2008 | 0.00 | 0.00 | 1.20 | 4.00 | 4.00 | 99.80 |
|       |          | 2014 | 0.00 | 0.00 | 1.00 | 3.19 | 3.00 | 99.90 |
| R5    | The closest distance to senior high school | 2008 | 0.00 | 1.00 | 3.90 | 8.38 | 8.00 | 99.00 |
|       |          | 2014 | 0.00 | 0.10 | 3.00 | 6.60 | 6.00 | 99.90 |
| R6    | The closest distance to the traditional market | 2008 | 0.00 | 1.00 | 3.00 | 10.11 | 9.00 | 99.80 |
|       |          | 2014 | 0.00 | 1.00 | 3.00 | 9.131 | 7.00 | 99.90 |
| R7    | The closest distance to the mall/shopping complex | 2008 | 0.00 | 2.00 | 6.00 | 17.19 | 18.00 | 99.80 |
|       |          | 2014 | 0.00 | 1.50 | 5.00 | 16.24 | 17.00 | 99.90 |
| R8    | The closest distance to cinema | 2008 | 0.00 | 34.40 | 98.00 | 68.42 | 98.00 | 99.80 |
|       |          | 2014 | 0.00 | 40.00 | 98.00 | 71.35 | 99.80 | 99.80 |
| R9    | The closest distance to hospital | 2008 | 0.00 | 7.60 | 19.00 | 29.62 | 41.00 | 99.80 |
|       |          | 2014 | 0.00 | 7.00 | 17.00 | 27.54 | 38.20 | 99.90 |
| R11   | Percentage of household with cable phone | 2008 | 0.00 | 0.00 | 0.00 | 4.393 | 1.90 | 100.00 |
|       |          | 2014 | 0.00 | 0.00 | 0.00 | 1.598 | 0.00 | 100.00 |
| R12   | Percentage of household with electricity | 2008 | 0.00 | 59.86 | 92.42 | 75.70 | 100.00 | 100.00 |
|       |          | 2014 | 0.00 | 94.68 | 100.00 | 90.83 | 100.00 | 100.00 |

Let see the mean and median values for each variable in 2008 and 2014. There are not any values of mean that equal with median values, this shows an asymmetrical distribution of data. Still in Table 7 where get IQR values from reduction Q3 and Q1. Long IQR values are found in variables R1, R2, R8, and R12. Long IQR shows that variable values are scattered or have a high distribution size. This varied distribution of data becomes difficult as a differentiator in determining urban or rural status.

The Results of the Kolmogorov Smirnov Test

Normally distributed data is considered to represent the population. Based on Figure 1 about the comparison of data distribution between 2008 and 2014, it is seen variables R1, R2, R11, and R12 have different patterns. But this assessment tends to be subjective, so it needs to be empirically tested.
The application of the K-S test on data in 2008 and 2014:

**Hypothesis:**
- \( H_0 \): dataset in 2008 and 2014 has the same distribution
- \( H_1 \): dataset in 2008 and 2014 has not the same distribution

**Level of significance:** \( \alpha = 0.05 \)

**Test Criteria:**
1. \( H_0 \) is rejected if \( D > D^*(\alpha) \)
2. \( H_0 \) is rejected if \( p\text{-value} < \alpha \)

In the 2008 and 2014 data set, there were 66,902 villages, therefore critical values for the K-S test \([D^*(\alpha)]\) i.e.:

\[
D_{n,m}(0.05) = 1.224 \times \sqrt{\frac{n+m}{nm}} = 1.224 \times \sqrt{\frac{66902+66902}{66902 \times 66902}} = 0.00669
\]

**Table 8.** Critical values of two-sample Kolmogorov Smirnov test, 2008 and 2014

| Index | Variable                                      | D     | p-value          | Decision     |
|-------|-----------------------------------------------|-------|-----------------|--------------|
| R1    | Population density                            | 0.01357 | 8.892e-06      | Rejected \( H_0 \) |
| R2    | Percentage of agricultural household          | 0.17851 | <2.2e-16        | Rejected \( H_0 \) |
| R3    | The closest distance to kindergarten           | 0.09693 | <2.2e-16        | Rejected \( H_0 \) |
| R4    | The closest distance to junior high school    | 0.08352 | <2.2e-16        | Rejected \( H_0 \) |
| R5    | The closest distance to senior high school    | 0.08731 | <2.2e-16        | Rejected \( H_0 \) |
| R6    | The closest distance to the traditional market| 0.03904 | <2.2e-16        | Rejected \( H_0 \) |
| R7    | The closest distance to the mall/shopping complex | 0.03339 | <2.2e-16        | Rejected \( H_0 \) |
| R8    | The closest distance to cinema                | 0.40404 | <2.2e-16        | Rejected \( H_0 \) |
| R9    | The closest distance to hospital              | 0.03723 | <2.2e-16        | Rejected \( H_0 \) |
| R11   | Percentage of household with cable phone      | 0.14908 | <2.2e-16        | Rejected \( H_0 \) |
| R12   | Percentage of household electricity           | 0.28280 | <2.2e-16        | Rejected \( H_0 \) |

From Table 8, it is known that all variables decided to Reject \( H_0 \), thus it is concluded that the data in 2008 and 2014 has not the same distribution. Based on D-value, the highest values are known in the variables i.e., the closest distance to the cinema, the percentage of households with electricity, percentage of agricultural households and the percentage of household with cable phone. This shows that it is necessary to modify or change the variables and criteria of the variables to produce indicators that represent actual urban or rural conditions.

4. **Conclusions**

Transition matrix of urban-rural status from applying 2010 scoring method on 2008 and 2014 data set shows that there are villages change from urban status to be rural status. The mean score of some variable is different inter-year. The data distributions of some variable are different inter-year. Based on the result of EDA above that it is reasonable to change the 2010 scoring method so that variables and criteria of each variable are appropriate to represent actual urban or rural conditions.

Variables that considering to be replaced, i.e. population density, the percentage of agricultural households, the existence of kindergarten, the existence of cinema, the existence of hotel/pub/beauty
shop, the percentage of household with cable phone and percentage of household with electricity. The criteria of all variables used in the 2010 scoring method need to be modified.

Advice for further research, i.e.: determination of appropriate variables as identification village according to urban-rural status and better criteria for each variable to be appropriate in determining rural-urban status.

References
[1][UN] United Nations. 2010. The Wye Group Handbook: Statistics on Rural Development and Agriculture Income (2nd Edition). United Nations.
[2][BPS] Statistics Indonesia. 2010. Peraturan Kepala Badan Pusat Statistik Nomor 37 Tahun 2010 tentang Klasifikasi Perkotaan dan Perdesaan di Indonesia. Jakarta (ID): BPS.
[3] Budiarto T. 2017. Kajian Tipologi Desa Berdasarkan Status Perkembangan dan Kemandirian Desa [thesis]. Bogor (ID): Institut Pertanian Bogor.
[4] Komorowski M, Marshall DC, Salciccioli JD and Crutain Y 2016 Exploratory Data Analysis in Secondary Analysis of Electronic Health Records (pp. 185–203). New York (US): Springer Publishing. https://doi.org/10.1007/978-3-319-43742-2_15.
[5] Saefuddin A, Notodiputro KA, Alamudi A and Sadik K 2009 Statistika Dasar (Jakarta (ID): Grasindo).