Poverty Data Model as Decision Tools in Planning Policy Development

Ahmad Haidar Mirza

Faculty of Computer Sciense, Universitas Bina Darma, Indonesia
Email: haidarmiza@mail.binadarma.ac.id

Abstract

Poverty is the main problem in a country both in developing countries to the developed countries, both in structural poverty, cultural and natural. That is, poverty is no longer seen as a measure of the failure of the Government to protect and fulfill the fundamental rights of its citizens but as a challenge of the nation to realize a fair society, prosperous and dignified sovereign. Various efforts have been made in determining government policy measures in an effort to overcome poverty, one of them by conducting a survey to assess the poor. The results of the survey of the various activities of the organization obtained a variety of database versions poverty to areas or locations. The information generated from the poverty database only includes recapitulation of poor people to the area or location. One step is to process the data on poverty in a process of Knowledge Discovery in Databases (KDD) to form a data mining poverty. Data mining is a logical combination of knowledge of data, and statistical analysis developed in the knowledge business or a process that uses statistical techniques, mathematics, artificial intelligence, artificial and machine-learning to extract and identify useful information for the relevant knowledge from various large databases.

Keywords: Poverty, Data Mining, KDD

1. INTRODUCTION

Poverty is a condition of a person who does not have income to meet the basic needs of food, and do not have the basic human ability to support sustainability. Later than the poverty according to UNDP is divided into four kinds, namely income poverty, human poverty, poverty and the basic needs poverty ability [1]. Poverty is the main problem in developing countries well in a Contracting State to the growing country, both structural poverty, cultural and natural. In developing countries poverty is the main job for government that must be addressed and should be resolved.

That is, poverty is no longer seen as a measure of the failure of the Government to protect and fulfill the fundamental rights of its citizens but as a challenge of a nation to realize a fair society, prosperous and dignified sovereign. Poverty has a negative impact which is spread (multiplier effects) against the social order as a whole. Various events of the conflict in the country that occur throughout the economic crisis, for example, reveals that the problem of poverty is not merely affect the economic resilience shown by the low purchasing power, but also affects the social resilience of communities and national defense. Many studies shows that poverty is also the estuary of other social problems. The problem of street children, acts of torture against children (child abuse), domestic violence, rundown houses, crime, alcoholism, ignorance, and unemployment is a result of poverty.

Poverty's definitions were complex, not only related to the economic dimension, but also with other dimensions beyond economics. But for this poverty is often conceptualized in the context of income and wealth insufficiency (lack of income and assets) to meet basic needs such as food, clothing, housing, education and health, which are within the scope of the economic dimension. Poverty is not only with respect to the level of income, but also of social, environmental and even empowerment and participation levels. Here are some definitions of poverty. Defines poverty as follows poverty is hunger [2]. Poverty is lack of shelter. Poverty is being sick and not being able to go to school and not knowing to know how to read. Poverty is not having a job, poverty is fear for the future, living one day at a time. Poverty is powerlessness, lack of representation and freedom.

At the World Conference on Social Development (the World Summit for Social Development) in Copenhagen in 1995, poverty is defined as follows poverty has a plural form, including low levels of income and productive resources that ensure a sustainable life; hunger and malnutrition; low levels of health; limitations and lack of access to education and other basic services; unnatural conditions and due to the disease continues to increase; mosey life and inadequate shelter; unsafe environment, as well as discrimination and social isolation; and also characterized by low levels of participation in the process of decision-making and in civil life, social and cultural rights.

Poverty is a failure for the functioning of some basic capabilities, or in other words a person is poor if the lack of opportunity to achieve/get this basic capability [3]. Poverty should not be considered only as a low income (low income), but should be regarded as the inability capability (capability handicap). According to the Central Statistics Agency (BPS) residents who are unable to meet the minimum basic needs of the population classified as poor. The value of the poverty line used is based on the minimum requirement of 2,100 kcal per capita per day plus the minimum requirement is a non-food basic needs someone who covers the basic needs for shelter, clothing, schools, transportation, as well as the needs of individual households and other basic. The amount of expenditure (in dollars) to meet the minimum basic needs of food and non-food poverty line is called [4]. Some of the poverty criteria established by other agencies, among others: BKKBN (National Family Planning Coordinating Board), the World Bank and UNDP (United Nations for Development Programs).

BKKBN define poverty based on criteria pre-prosperous family (pre-KS) and family welfare I (KS I). World Bank set of poverty based on income per person per day. Usually size is US $ 1 or US $ 2. Residents with income below the nominal value are categorized as poor. UNDP in the 1990s introduced the human development index (human development index - HDI) and human poverty index (human poverty index - HPI). Compared with the World Bank poverty criteria, the UNDP relatively more comprehensive approach. UNDP approach includes not only the economic aspects (income), but also education (literacy rate) and health (life expectancy). Calculation method of measuring poverty Poverty conducted CBS since the first time to the present using the same approach that the basic needs approach (basic needs approach). With this approach, poverty is defined as the inability to meet basic needs.

In other words, poverty is seen as an economic inability to meet the needs of food and non-food fundamental.

Based on the Regulation of the Ministry of Internal Affairs No. 42 of 2010 on the Provincial Poverty Reduction Coordination Team and District/City, Article 1 refers to poverty alleviation programs are activities carried out by the government, local governments, businesses, and communities to improve the welfare of the poor through social assistance, community empowerment, economic empowerment of micro and small enterprises, as well as other programs in order to increase economic activity. For the district level Regent/Mayor is responsible for accelerating the implementation of poverty alleviation in the district/city.

Various efforts have been made in determining government policy measures in an effort to overcome poverty, one of them by conducting a survey to assess the poor. This step is done by the agency or organization that requires data poor to implement improvement programs welfare of the people. Even recently the National Movement of Poverty Reduction Team (TNP2K) launched the Integrated Data Base website for Social rock Program, with the aim that the Social Assistance Program can be implemented appropriately in accordance with the participation and program objectives. Results of surveys of various organizations obtained various versions of the poverty database for area or location. The information generated from the poverty database only includes recapitulation of poor people to the area or location.

One step to process the data on poverty is using a process of Knowledge Discovery in Databases (KDD) to form a data mining poverty. Data mining is a logical combination of knowledge of data, and statistical analysis developed in the knowledge business or a process that uses statistical techniques, mathematics, artificial intelligence, artificial and machine-learning to extract and identify useful information for the relevant knowledge from various databases large. This process is intended to undergo extraction poverty information in the database in order to get a pattern of knowledge-based information that can be used in policy-making remains within the target to determine the pace of poverty reduction [5].

2.   METHODS
2.1. Research Method
In conducting this research, method used is the method of Discovey Knowledge in Databases (KDD). Knowledge Discovery In Databases (KDD), is an activity that includes the collection, use historical data to determine regularities, patterns or relationships in a large data set [6]. The output of data mining is widely used for decision making in the future. An overview of the KDD process can be seen as Figure 1.
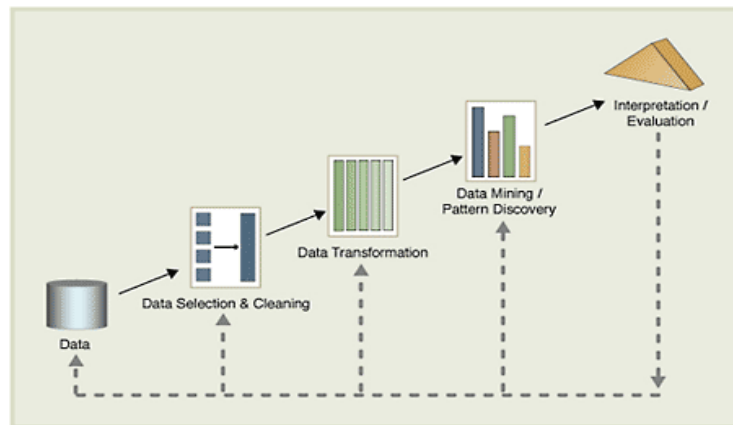
Figure 1. *Knowlarge Discovery in Database*

From the figure above shows that the KDD process consists of:
1. Data Understanding, which is the process of understanding the data based on the needs of the data (Data Requirement). This process includes collecting data (initial collection of data) and the description of the data (data decription).
2. Data Preparation, namely preprocessing consisting of Selection Data (Data Selection) and cleaning (cleaning) of data, the data selection process is carried out that are tailored to the needs and cleansing data from its data redundancy or data with the wrong type of data .
3. Data transformation, the process of converting data into another format in accordance with the needs of analysis.
4. Modeling data mining, which is a process to obtain the pattern and characteristics of the data, in this phase used clustering methods.

Interpretation/Evaluation, interpretation and evaluation of the problems faced by the data analyzed.

3. RESULTS AND DISCUSSION
Research carried out following the steps in the method of Knowledge Discovery in Databases (KDD).

3.1. Data Understanding
3.1.1. Data Requirement
In order for an understanding of data that can be done right on target, the researchers conducted observations to government authorities and is responsible for the direction of the development plan will be implemented government annually. The planning proposal came from people who channeled through muresbang subdistricts and proposals from government agencies/SKPDs as public works Department, Social Services, etc. The problem of poverty is not a new problem faced by the government from year to year. This is related to the development plan will be implemented primarily concerned with the development of infrastructure such as schools, roads, bridges, proposition clean water, electricity and others. Community development

projects to be encouraged, not just a distribution of temporary cash assistance and tends not too optimal in poverty alleviation.

From the above explanation, the researcher is interested to see the connection pattern of the datas per region on poverty by age, level of education, employment, housing conditions, how the acquisition of drinking water and lighting which is an indicator of poverty that is applied by the BPS.

3.1.2. Initial Data Collection
This process is a process of data collection in which data is collected and used comes from poverty database PPLS designed by BPS cooperate with TNP2K in 2011 in the district Ogan Komering Ilir (OKI) in the province of South Sumatra. The data used include family data and poverty data from 18 sub-districts in Ogan Komering Ilir, South Sumatra province.

3.1.3. Data Decription
Is the process whereby data description family data include family relationship to head of household, age, sex, marital status, physical disability, school, diploma, and work. While the data consists of attribute address poverty, and poverty indicators and poverty class (Almost Poor/Poor/very Poor). Poverty's indicators can be grouped into several groups, such as: conditions of housing, clean water source, lighting/electrical, fuel used, landfills, vehicle ownership and participant in assistance program (PKH & Raskin).

3.2. Preparation Data
3.2.1. Data Selection
The data used in this study include family data record number of 269 590 (Table a1602) and poverty data as much as 78 684 records (Table a1602). In addition there are several master table that stores information subdistrict and village names.

3.2.2. Cleaning Data
Phase of data cleaning is the initial phase of the KDD process. At this stage the data that are irrelevant, missingvalue, and redundant should be in the clear. This is because the relevant data, not missingvalue, and not redundant is the initial terms of doing data mining. A missing value if the data is said to be contained in the dataset attributes that do not contain or empty value, while the data is said to be redundant if in a dataset contained more than one record that contains the same value.

3.3. *Data Transformation*
Stages of data transformation is a phase change data into a form suitable for mining. At this stage, the merger/integration prior to the several data sources that include family data, poverty data, the data sub-district and village data into an integration of data that is adjusted to the target data to be used by the data mining process. As described in Figure 2.
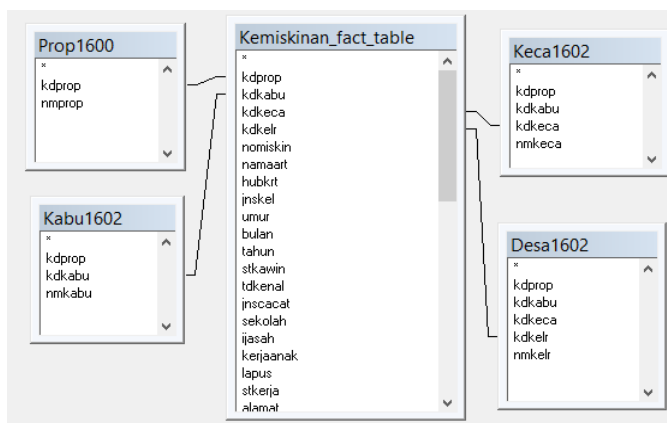
Figure 2. Data Transformation

In Figure 2 illustrates the process of merging the two data sources from a database PPLS is tables of family (a1602) and tables of poverty (r1602). The process of merging the two tables to form a table FACT which is poverty table fact. The table will be related to the provincial table (prop1600), tables district (kabu1602), sub-district table (keca1602) and tables village (desa1602) as a dimension table that will be used in the data mining process.

3.4. Data Mining
Stages of data mining is the stage to obtain the information hidden in the database. Process Data mining is the process of searching for a pattern or interesting information in the selected data by using techniques or methods. Technique, method, or algorithm in data mining is very varied. Selection of appropriate methods or algorithms are very dependent on the purpose and overall KDD process. Selection of the poverty goal of data mining is clustering which purposed to group data based poverty with similar characteristics to the same area and the data with different characteristics to other areas. In this clustering stage grouping data grouped by grouping data based on region of residence of the poor which includes groups district, village and rural/neighborhood and indicators of poverty. Based on the needs of the data (Data Requirements) has been done in previous stages of the importance of data mining models of poverty as seen in the Table 1.

Table 1. Model of Poverty Data Mining

Graphic	Definition	Patterns of Information
Graphic	Definition	Information on the distribution of poor information for each region
Graphic	Definition	Information on the level of education of the population and poverty rates by region
Graphic	Definition	Information age level (children, productive, non productive) with poverty level per region
Graphic	Definition	Information on living conditions with poverty per region

Graphic	Definition	Information on how to obtain drinking water and lighting with poverty per region
Graphic	Definition	Job information with poverty levels per region
Graphic	Definition	Information of gender with poverty rates by region

Data mining models generated is displayed in the form of dimensional data and graphs. There are seven patterns of information that is displayed in accordance with the needs of the data that has been defined previously.

1. Distribution of poor information for each region.
2. The level of education of the population and poverty rates by region.
3. The age of the population and poverty rates by region.
4. Living conditions and poverty per region.
5. How the acquisition of drinking water and lighting with poverty rates by region.
6. Job with poverty rates by region.
7. Gender with poverty rates by region.

3.4.1. Poor Dissemination of Information for Each Region

Cluster/poor population groups formed by area and population poverty level. To illustrate the conditions of the poor for each region based on the level of poverty population (HM, M, SM). The spread of poverty in each of the districts, even to rural villages can be seen based on the data model are visualized in graphs and tables multidimensional data can be seen in Figure 3.
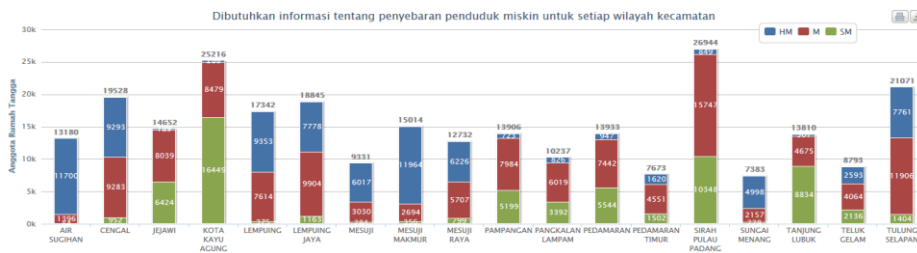

Figure 3. Graphs of Poor Spread Per Subdistrict

3.4.2. The level of education of the population and poverty rates by region

Cluster/poor population groups are formed based on education and poverty population. The level of poverty which are grouped into several levels: 0 - no school, 1 - SD, 2 - Junior, 3 - high school, 4 -College. Data model of poverty and education level can be visualized with the following chart in Figure 4.

Figure 4. Graphs of Poverty and Education Level

Based on the above data model then the government can see the state of education of the population through poverty's indicators such as school enrollment rate (APS), early childhood education, educational facilities, the availability of teachers, the ratio of the availability of school, the student/teacher. In addition, the higher education level, the high needs of cost, it means not all residents are able to reach. In this case government needs to establish school infrastructure, increase the number of teachers, providing assistance/scholarships or promoting free education program. In order to do so would require accurate information about the level of education per region.

3.4.3. Age population and Poverty Rate per Region

Cluster/poor population groups are formed based on age and level of poverty of the population. The age of the population classified by groups: children, ages productive and non-productive age. The resulting data model can be visualized in graphical form in Figure 5.
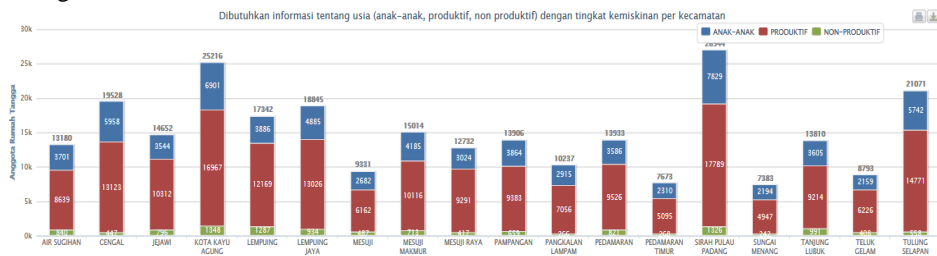

Figure 5. Graphs of Age Population and Poverty Rate

3.4.4. Living condition and Poverty Rate per Region

Cluster/group formed by Events poor people residence. The living conditions in the population can be recorded with the type of flooring and walls of houses. Good home must meet eligibility requirements as a residence. Some variables are often used to assess the feasibility of residence is the floor, walls and roof. The better the quality of the floor, walls and roof, reflecting the high socio-economic conditions. In 2011 collection PPLS floor types are grouped into 3 types, not soil/bamboo, soil and bamboo. It shows the higher level of welfare of a household, the higher the quality of the type of floor. Instead of households that have the type of bamboo flooring from declining percentage as the higher level of welfare. Graphs of livng condition and poverty rate can be seen in Figure 6.
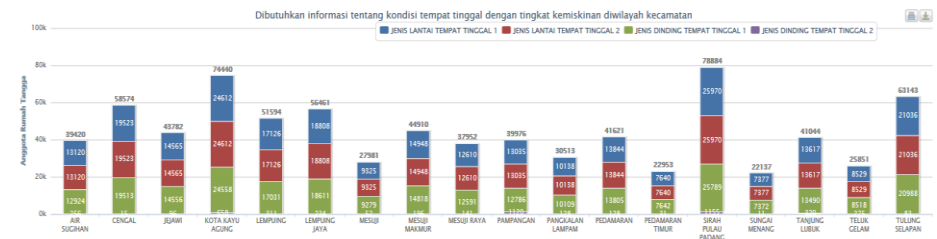

Figure 6. Graphs of Livng condition and Poverty Rate

According to the existing BPS 14 fourteen criteria poor families, among other residential building floor area of less than 8 m2 per person, the type of floor residential buildings made of soil/bamboo/wood cheap, kind of wall shelter made of bamboo/thatch/low quality of wood/wall without plaster, do not have the facilities to defecate/together with other households, household lighting source does not use electricity. Through the poverty criteria, there are many families in Indonesia are categorized as below the poverty line, pre-prosperous families, poor families and designation criteria poor. By those criterias, the government can provide appropriate housing and rehabilitation to place them and to solve problems triggers associated with poverty. One of the government programs is home assistance program for the poor, named Stimulants Help Housing Organization (BSPs). The program has launched the Ministry of Housing (Kemenpera) RI. In this case, nationally Kemenpera targeting disbursements BSPs for approximately 250,000 housing units with a value of Rp 1.6 trillion. However, the distribution is deemed not home right on target because there are many poor people who do not get the housing assistance. This is due to the lack of information to the government about the data of poor people who have a house uninhabitable. Data mining models generated can categorize the poor by region and their living conditions of the district, village up to the countryside. Here's an example display multidimensional data model for conditions where the level of poverty per region, district, village and countryside

3.4.5. Way Acquisition of Drinking Water and Lighting with Poverty Rates by Region

Cluster/poor population groups are formed by way of acquisition of drinking water and lighting. The acquisition of drinking water were divided into three groups: water source 1 = wells, water source 2 = rivers, water source 3 = rain. While the lighting consists of 3 groups: 1 = electric lighting, 2 = non-electric lighting. Graphs of water source condition, lighting and poverty rates can be seen in Figure 7.
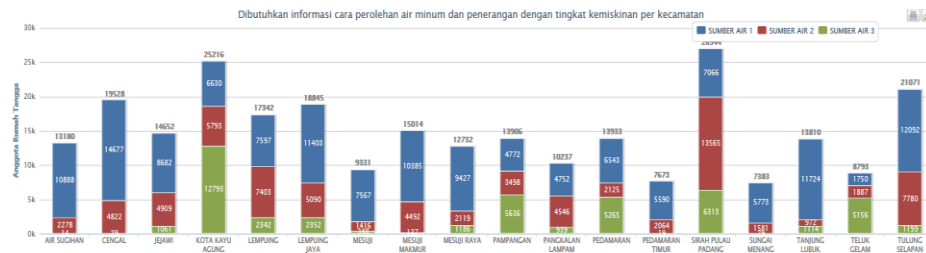

Figure 7. Graphs of Water Source Condition, Lighting and Poverty Rates.

Source of drinking water that is used to describe the level of welfare in the field of health. So that water is used for drinking water should be clean. Clean water is one of the types of resources of good quality water-based and commonly used by humans for consumption or in performing their daily activities. Clean water is strongly influenced by the source of drinking water. The further away from the source of pollution, then the water will be clean and healthy to consume. Clean water and decent to drink is water sourced, such as bottled water, refill water, tap, boreholes, protected wells, protected springs. According to data PPLS 2011, almost 33.06% of the houses date in Indonesia using wells as a source of clean water.

According to the results of the Village Potential Data Collection (Podes) in 2011, more than 96 percent of the village the majority of households use electricity as the main source of lighting, both electric and non electric. This is consistent with the results of PPLS 2011, that the majority of households are using electricity, both electric and non-electric. Government assistance programs for the poor, especially for the provision of clean water and lighting is one of the main things that must be met to improve the health of people, especially the poor in rural areas. Data mining models generated formed clusters/poor grouping based sources of drinking water, lighting and per region population poverty level. With this model the government can find clusters/groups of poor people who needs social assistance and infrastructure development in accordance with the conditions of the people and groups in their territory. This model can classified poor people according to how they get the lighting and drinking water per district, village, rural.

3.4.6. Job and Poverty Rate by Region

Cluster/poor population groups are formed based on the employment status of the population. The job of the population can be classified into: 1-Trying own, 2-Employer assisted by temporary workers/not paid, 3-Employer assisted by permanent workers/paid, 4-workers/employees/private, 5-Free Workers, Workers 6-family/not paid. Graphs of statue on job and poverty rate can be seen in Figure 8.
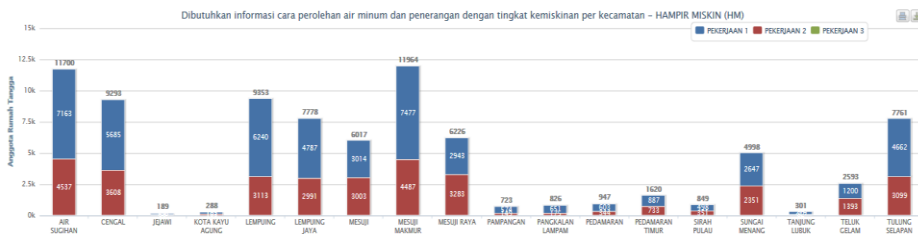

Figure 8. Graphs of Statue on Job and Poverty Rate

Main employment status of population aged 15 years and older showed a different phenomenon from each household group. When viewed from the data largely existed as family workers, laborers/employees and workers are free, while the other group mostly worked as laborers/employees. When observed difference is due to differences in education levels. In population groups of workers/employees have a relatively higher level of education, so have a greater opportunity to work in the formal sector as laborers/employees.

3.4.7. Gender Population and Poverty Rate by Region

Cluster/poor population groups are formed based on the gender of the population. Graphs of gender and poverty rates can be seen in Figure 9.

Figure 9. Graphs of Gender and Poverty Rates.

From the resulting data mining models obtained that clusters/groups no significant difference of gender-sex groups of men or women in all regions. It is seen from the spread of the poor in sub-districts, villages and countryside of each gender does not have a large difference between the population gender male or female. It can provide input to the government to focus more on overall poverty reduction programs without having constrained problem of gender/sex of the poor.

3.5. Interpretation/Evaluation

Information patterned generated from the data mining process needs to be displayed in a form that is easily understood by interested parties. This stage is part of the KDD process called interpretation. This phase includes checking whether a pattern or information found in conflict with the fact that there was previously. From the results of data mining stages in seven patterns of information can be generated in accordance with the results of the analysis at this stage is done on the data requirements by Badan Perencanaan dan Pendapatan daera (BAPPEDA) in the district Ogan Komering Ilir and Banyuasin district in the province of South Sumatra. Furthermore, data mining models will be transformed into an application form that will be easily accessible globally by the relevant government authorities.

4. CONCLUSION

Based on the research that has been done then obtained seven patterns of information on the results of KDD methods that can be used to assist the government in understanding the patterns of poverty data. Seventh information pattern is the result of grouping poverty data by region and level of poverty based on poverty indicators distribution of poor information for each region, the level of education of the population and poverty rates by region, the age of the population and poverty rates by region, living conditions and poverty per region, how the acquisition of drinking water and lighting with poverty rates by region, job with poverty rates by region, and gender with poverty rates by region. The results of the seven patterns on this information will be transformed into the form of a web-based application that can be used by the government in understanding poverty data patterns. It helpful to provide useful information for the government in making decision on development planning policy which will impact on reduction of poverty.

5. REFERENCES

[1] UNDP. (2011, April 15). Definisi Kemiskinan Menurut United Nations for Development Programs. Retrieved April 26, 2013, from SCRIBD:

http://www.scribd.com/docl/44466968/Pembangunan-dan-Kemiskinan   Perempuan -di-Kota

[2] World Bank. 2000. World Development Report 2000/2001: Attacking Poverty, Washington, DC: World Bank

[3]   Sen, 1985, *Inequality Reexamined*. Harvard University Press.

[4]   BPS. 2007. Berita Resmi Statistik, No. 38/07/Th. X, 2 Juli 2007

[5]   Therling K. 2006. An Introduction to DataMining: Discovering hidden value in your data warehouse, www.thearling.com, diakses tanggal 2 April 2013

[6]   Santoso, Budi., 2007, Data Mining Teknis Pemanfaatan Data untuk Keperluan Bisnis, Graha Ilmu.