G306

LEVEL PREDICTION OF PNEUMONIA IN CHILDREN TIME RANGE BASED BREASTFEEDING, IMMUNIZATION STATUS, NUTRITION USING K-MEANS

Yusriel Ardian
Faculty of Information Technology, University of Malang
Kanjuruhan
acilnet@yahoo.com

Wiji Setiyaningsih
Faculty of Information Technology, University of Malang
Kanjuruhan
wiji_setiya@yahoo.co.id

Abstract - At the core contribution of which can be obtained from the results of this study are related parties of the neighborhood health center level, health centers and hospitals, particularly hospitals that become the object of study can obtain information that supported accurate data about the importance of exclusive breastfeeding, immunization and nutrition in toddlers, in order to suppress or minimize disease of children under five will be. Related parties as early as possible to disseminate to the public the relationship between exclusive breastfeeding, immunization and nutritional status on disease susceptibility that can be suffered by a toddler.NWith the implementation of K - Means algorithm in the process of knowledge about high clusterisasi pneumonia in toddlers based immunization status, the time range of exclusive breastfeeding, and nutrition status of children under five, are expected to contribute grouping and determination of the most appropriate number of clusters/accurate to predict the value of level of frequency/frequency of visits toddler to the hospital for treatment . The proposed plan of activities to collect and study the literature relating to the concept of DM clustering, which uses K-Means algorithm . Literature sources such as text books, papers, journals, scientific papers.

Keywords : clustering, K-Means, Pneumonia, ISPA, nutrition, immunization, exclusive breastfeeding.

I. PRELIMINARY

ISPA is a disease affecting one or more of the parts and the respiratory tract from the nose (upper line) to the alveoli (bottom line) including adneksanya networks, such as the sinuses, middle ear and pleural cavity. While the infection that attacks the lower respiratory tract (lungs) one of which is pneumonia. Pneumonia is an acute inflammatory process in the lung tissue (alveoli) due to infectious germs that cause respiratory distress. Pneumonia is dangerous because it can lead to death, because the lungs are not able to function to get oxygen to the body (Depkes RI, 2007). Pneumonia has become a health problem in the world because of the high death rate. This does not just happen in developing countries, but also in developed countries.

According to research conducted Mery and Widyaiswara (2012), which attacked pneumonia in infants is influenced by the mother's level of education, immunization status, exclusive breastfeeding, and nutritional status of children under five. Meanwhile, according to Dea and Yandofa (2012) found a significant association between nutritional status on the incidence of pneumonia in young children, and there is a significant association between breast feeding on the incidence of pneumonia in young children.

The World Health Organization (WHO) in 2005 estimated deaths from pneumonia in children under five worldwide about 19 percent or about 1.6 to 2.2 million. Of which about 70 percent occur in developing countries, especially Africa and Southeast Asia. World Pneumonia Day (WPD) reported that Indonesia is the country with the incidence of pneumonia 6th rank in the world. In Indonesia, pneumonia is the third cause of death after cardiovascular and tuberculosis. Pneumonia in children under five mortality rate in Indonesia is estimated at 21%. The morbidity rate was estimated at 250 to 299 per 1,000 children under five each year.

Dr. I. Boediman, Sp. A (K) in the World Pneumonia Day 2010 seminar revealed that the child has a healthy immune system that protects the lungs from germs. Child with a weak immune system such as child malnutrition, especially because not exclusively breastfed, vitamin A deficiency and measles has a high risk of pneumonia (Sutriyanto, 2011). The high rates of respiratory tract infections in infants related to environmental sanitation, inadequate health services and with low immunization coverage. Respiratory tract infections in infants are also influenced by the pattern of breastfeeding and complementary feeding. In infants who had been given food before the age of 4-6 months or even a few moments after

birth can lead to infants susceptible to disease infection (LIPI, 2004).

According to Professor Guido Moro Macedonis Melloni Maternity Hospital of Milan about two- thirds of the benefits of breastfeeding the baby's immune system is in the stomach , so it is important to pay attention to what the baby is eating and drinking . That is why baby mothers breastfed newborn desperately need especially during the first 6 months of life.

As the baby 's first food , it turns out not only nutrients breast milk is perfect for baby and closer emotional relationship between mother and baby, but at the same time provide protection for beneficial breastfeeding strengthens the natural immunity of newborns . So many benefits of breast milk for the baby, ten magic among other things: 1). Breast milk strengthens the immune system. The main components of the immune system builder ASI is a prebiotic. 2). Breastfeeding lowers the risk of allergies. 3). Breastfeeding lowers the risk of gastrointestinal disease, such as diarrhea and improve immunity in the digestive system. 4). Breastfeeding lowers the risk of respiratory problems , such as colds and coughs. 5). Breast milk is rich in AA | DHA that support the growth of children's intelligence. 6). Breast milk contains natural prebiotics to support the growth of intestinal flora. 7). Breast milk composition proper nutrition and a balanced (where asi only have one).

Health Ministry data shows, breast milk can reduce mortality by 17 percent in the new birth (neonatal) and 12 per cent in children under five years old. Newborn mortality rate nationally is 34 per 1,000 live births and the mortality rate of children under five years old to 44 children per 1,000 live births. Regrettably, the rate of exclusive breastfeeding in the country is still very low. Only about 22 percent of mothers giving birth to her baby exclusively breastfed. The results of recent research investigators showed that infants who received complementary foods before 6 months of age (non- exclusive breastfeeding) would be more frequent diarrhea, constipation, respiratory infection (Soraya , 2005) (Suprihadi, 2010).

Based on the description of the literature review studies mentioned above, the knowledge of the high pneumonia in toddlers is an important thing that needs to be investigated, particularly by common factors known to mothers who have children under five as immunization status, time range of exclusive breastfeeding, and nutritional status in toddlers.

There is a field of science that is able to solve the problem is data mining. Data mining is an activity that includes the collection, use historical data to determine the regularities, patterns or relationships in large data sets. One of the major tasks of data mining is the clustering of clustering where the data is grouped not have any examples of groups (Santosa, 2007).

One clustering method which can be used as a solution to these problems is the method of K - Means, as was done by the researcher Dwi Noviati (2012) for regional revenue budget, or Budi and Patdono for market segmentation.

With the implementation of the K-Means algorithm in the process of knowledge about high clusterisasi pneumonia in toddlers based immunization status, time range of exclusive breastfeeding, and nutritional status of children under five, are expected to contribute grouping and the determination of the most appropriate number of clusters/accurate to predict the value of level of frequency/frequency of visits toddler to hospital for treatment. In essence the contribution that can be obtained from the results of this study are related parties of the neighborhood health center level, health centers and hospitals, particularly hospitals that become the object of study can obtain information that supported accurate data about the importance of exclusive breastfeeding, immunization and nutrition in infants, in order to suppress or minimize disease of children under five will be. Related parties as early as possible to disseminate to the public the relationship between exclusive breastfeeding, immunization and nutritional status on disease susceptibility that can be suffered by toddlers.

II. LITERATURE NUTRITIONAL STATUS AND THE BREASTFEEDING TODDLERS

Many factors affect the high number of deaths from pneumonia in children under five, namely: age, gender, malnutrition, low birth weight, breastfeeding status, immunization status, kepadatanhunian, ventilation, air pollution in the home. The high incidence of pneumonia primarily affects infants and toddlers age group. Toddlers with poor nutrition will be more susceptible to pneumonia than infants with normal nutrition for endurance factor is less. Infectious diseases alone would cause infants do not have the appetite and lead to malnutrition. In a state of malnutrition, infants are more susceptible to pneumonia attack even longer (Dea Yandofa, 2012).

One of the risk factors that play a role in the incidence of pneumonia in children is nutritional status, where the interaction between infection and protein-energy malnutrition (PEM) has long been recognized, both conditions synergistic, mutual influence, which predisposes to the other one. In PEM, decreased body resistance and virulence of pathogens causing stronger balance will be disturbed and infection, whereas one of the main determinants in maintaining that balance is the nutritional status of children . Immune system in infants or toddlers not yet completely formed. Therefore, the baby will be more susceptible to infections if they do not get adequate nutrition. This is reinforced by research Rusepno (2005) which says that nutrition and infection are factors that affect the growth of children in developing countries, including Indonesia. Plus the delay in providing nutrition services will result in damage that is difficult and may not even be helped (Dea Yandofa, 2012).

III. RESEARCH METHODS

Implementation methods of research used in writing this time is the experimental method. Experimental method is a

design study that identified a causal relationship (Sudaryono et al: 45, 2011).

The explanation of this research method is as follows:

A. Literature Study

By collecting and studying the literature relating to the concept of DM clustering, which uses K-Means algorithm. Literature sources such as text books, papers, journal, scientific papers, and supporting websites.

B. Archive Data Collection

To find the necessary information, the author conducted data collection archives.

C. Preprocessing Of Data

Preprocessing the data include:

• Selection of data

To select a data set (dataset) which will be used in this paper, namely the data range time exclusive breastfeeding, immunization status, and nutrition.

• Cleaning

To clean the data, the complete data, removing duplicate data, remove noise.

• Transformation of data

To format the data to be ready clustered.

D. Clustering Using K-Means Algorithm

Stage of the process where data is already dipraproses in the cluster by using the workings of K-Means algorithm.

• Select the number of clusters k. Initialize k cluster centers can be done in various ways. How random is often used, the cluster centers are given by the initial value and the random numbers are used as initial cluster centers.

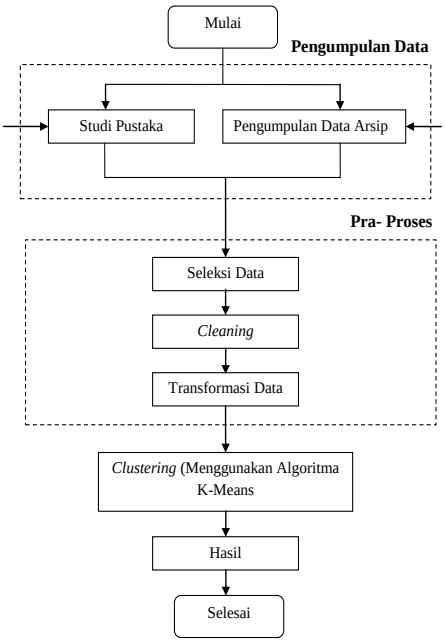

Figure 1. Stages of Research

E. Place each data/object to the nearest cluster, proximity of the two objects is determined by the distance between the two objects . Similarly, the proximity of the data to a particular cluster is determined by the distance between the data center of the cluster. In this stage, the data need to be calculated for each distance to each cluster center. The closest distance between the data at a particular cluster will determine the data included in the cluster where. As for the distance calculation using the formula Eulidean.

$d\ (x\ ,\ y) = \sqrt{}\ (xi - yi)^2 + (xi - yi)^2$ ..... (1)

• Recalculate the center of the cluster with cluster membership now . Center of the cluster is the average of all the data/objects within a particular cluster. Calculations by determining the centroid/center of the cluster.

$C\ (i) = x1 + x2 + x.. + x... \Sigma x$ ..... (2)

• The shortest distance between the center of the cluster with the data/object determines the position of a cluster of data/objects. For example, a data/object A has the shortest distance to the cluster center compared to one another, then the data/object A went into cluster 1.

• Reassign each object by using the new cluster centers. If the center of the cluster is no longer changed, then the process is complete pengclusteran. If changed, then go back to step no.3 to the cluster center does not change anymore.

• Once the process is complete pengclusteran, it will be calculated the value of each cluster SSE. SSE value depends on the number of clusters and how the data are grouped into clusters such. The goal is to obtain a partition (fixed number of clusters) which minimizes the total square error. The smaller the value of SSE, the better the clustering results. Here's how the SSE.

$SSE = (Ci)^2 + (Ci)^2 + (C..)^2 + (C..)^2$ ..... (3)

F. Analysis Of The Results Clusterisasi

Stages to analyze the results already obtained in the clustering process.

IV. DISCUSSION

A. Aspects of Assessment

Aspects value obtained from the City Health Office in Malang, where its aspects, among others:

a. Exclusive breastfeeding status

Exclusive status data contains values YES and NO

a. Immunization Status

Exclusive status data contains values YES and NO

c. Nutritional intake

Exclusive status data contains values and LESS GOOD

B. System Design

Context Diagram

The following is a general overview of the system K-Means method for Disease Pneumonia high Knowing Striking Toddler Time Range Based on exclusive breastfeeding, immunization and nutrition status in the context diagram form.

In general the system can be described that the admin role is to provide input to the system in the form of data on the number of pneumonia patients based on three aspects of previous years (the value for each cluster), which serves as the training data. Next the user can provide input the number of patients by 1 symptom/cluster to predict the amount based on other symptoms. The results of prediction are also able to demonstrate the prediction of the overall rate of patients with pneumonia symptoms/clusters.

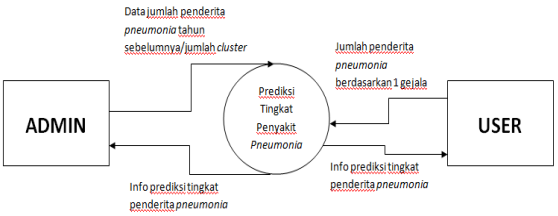

Figure 2. Prediction Context Level Diagram of Pneumonia Patients

C. Data Flow Diagrams (DFD Level I)

The following description of the system K-Mean Method For High Knowing Pneumonia Patients Striking Toddler Time Range Based on exclusive breastfeeding, immunization and nutrition status in the form of Data Flow Diagrams (DFD Level I).

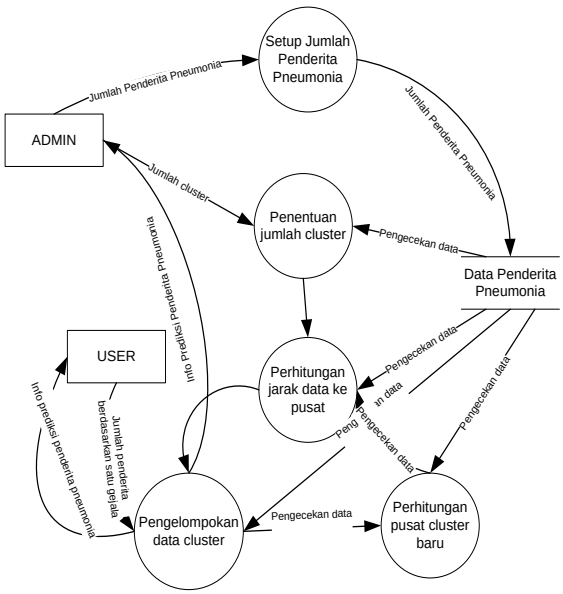

Figure 3. Data Flow Diagrams (DFD Level I) Predicted Rate Pneumonia Patients

In DFD level I explain step-step completion prediction using K-Means method. Starting from the master setup process training data pneumonia number of people with previous years (the value for each cluster), so that the training data are stored in files. Next training data used for the calculation of the K-Means method, starting from the centroid calculation, calculation of distances, to obtain the value of sum square error (SSE).

D. Database Design

Application of K-Means method to determine the height of pneumonia in toddlers is based on data summary contracted pneumonia caused due to problems of time range of exclusive breastfeeding, immunization status, and nutritional intake needed to use one table as training data for the prediction of future data beracuan six last year . Adapaun design training data tables are as follows:

TABLE 1. TRAINING DATA TABLE PNEUMONIA

Nama Field	Tipe Data	Keterangan
code Year	Byte	Code as the reference year data summary per year
breast milk	Integer	Number of people per year who contracted pneumonia as a result of problems on exclusive breastfeeding
immunization	Integer	Number of people per year who contracted pneumonia as a result of immunization issues
Nutritional intake	Integer	Number of people per year who contracted pneumonia as a result of problems with the intake of nutrients

E. Discussion on Application of K-Means Method

In this study using three aspects of the issue on breastfeeding, immunization, and nutrition intake, with the last six years of training data as follows:

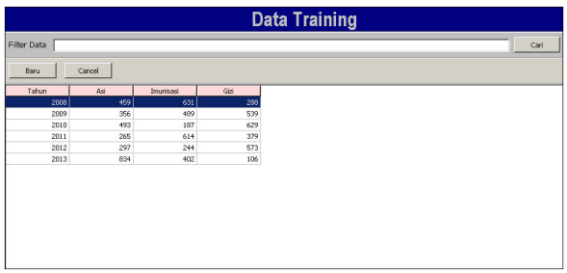

Figure 4. Training Data Pneumonia Patients Per Year

In this stage the researcher determines the value of the centroid into two parts, ie having 3 and 2 centroid centroid. This is done to determine where the most amount of good centroid, as in the formulation of the K-Means cluster most smallest value will be the value of the most well clusters. Here is the process of calculating the application of K-Means method to predict the rate of pneumonia disease that strikes children under five years in the future in Malang, due to the

impact of the frequency of exclusive breastfeeding, immunization status and nutritional intake, with beracuan on public health training data the last six years.

F. Initial Centroid Calculation

Clusterisasi phases using K-Means algorithm, beginning with the formation of clusters in a dataset are 3 clusters (exclusive breastfeeding, immunization, and nutrition intake), with 2 parameter testing, namely, 3 and 2 values centroid centroid value.

Perhitungan Centroid Awal

	407.5	379.0	565.5	436.0	465.33
	560.0	400.5	323.0	435.67	420.0
	413.5	504.0	339.5	485.33	352.67

Figure 5. Initial Centroid Calculation

3 centroid calculation to obtain C0, C1, and C2 for each cluster. 3 For the calculation of the centroid is obtained from the sum of two values for each aspect of the training data is divided by 2. As for the centroid obtained 2 C0, and C1, which is obtained from the sum of three values for each aspect of the training data is divided by 3.

Calculation of Distance Early

perhitungan Jarak Awal

	153.11	325.86	329.94	278.61	220.78
	153.11	97.91	333.54	110.11	226.79
	439.18	272.4	327.97	292.79	362.51
	156.24	272.4	420.17	268.98	280.11
	370.82	189.68	364.49	252.47	328.42
	549.02	604.51	364.49	550.85	443.94

Figure 6. Calculation of Distance Early

In this step the distance calculations carried out to determine the distance of each outcome data on the number of clusters (k) at each centroid. Perform calculations to determine the distance of each data with initial centroid, using the formula euclidiance distance. Here are the results for 3 centroid distance calculation, and 2 centroid.

Centroid Calculation for Iteration 1

Based on how the K-Means algorithm as specified value clusters (k) and then calculate the distance between the centroid and data on each of each centroid. In this stage, the calculation returns a value to the centroid of each cluster, called iterations, until the value has not changed from the previous centroid.
3 centroid calculation to obtain C0, C1, and C2 for each cluster. 3 For the calculation of the centroid is obtained from the sum of two values for each aspect of the training data is divided by 2. As for the C0 2 centroid obtained from the sum of two values for each aspect of the training data is divided by 2, and C1 were obtained from the sum of 4 values for each aspect of the training data is divided by 4.

Perhitungan Centroid pada Iterasi 1

	407.5	379.0	565.5	407.5	472.25
	560.0	400.5	323.0	560.0	361.75
	413.5	504.0	339.5	413.5	421.75

Figure 7. Centroid Calculation for Iteration 1

3 centroid value for iteration 1 to step 3, the results remained similar to 3 centroid in step 1, then the iteration is stopped.
Centroid value for iteration 2 to 1 step 3, with the result 2 DIFFERENT centroid in step 1, then the second calculation.

Distance calculation for Iteration 1

Distance calculations for iteration 1 is done, because the value of 2 centroid step 3, with the result DIFFERENT 2 centroid value in step 1. Perform calculations to determine the distance of each data with initial centroid, using the formula euclidiance distance. Here are the results for 3 centroid distance calculation, and 2 centroid.

Perhitungan Jarak pada Iterasi 1

	153.11	325.86	329.94	153.11	300.93
	153.11	97.91	333.54	153.11	208.46
	439.18	272.4	327.97	439.18	271.88
	156.24	272.4	420.17	156.24	329.26
	370.82	189.68	364.49	370.82	259.72
	549.02	604.51	364.49	549.02	481.85

Figure 8. Distance calculation for Iteration 1

Centroid Calculation for 2nd Iteration

3 centroid calculation to obtain C0, C1, and C2 for each cluster. 3 For the calculation of the centroid is obtained from the sum of two values for each aspect of the training data is divided by 2. As for the C0 2 centroid obtained from the sum of two values for each aspect of the training data is divided by 2, and C1 were obtained from the sum of 4 values for each aspect of the training data is divided by 4. In the following table the result of the calculation of the centroid for iteration 2.

Perhitungan Centroid pada Iterasi 2

	407.5	379.0	565.5	407.5	472.25
	560.0	400.5	323.0	560.0	361.75
	413.5	504.0	339.5	413.5	421.75

Figure 9. Centroid calculations for iteration 2

3 centroid value for the 2nd iteration step 5, the result is still the same with 3 centroid in step 3, then the iteration is stopped.

Centroid value for iteration 2 to 2 step 5, the result is still the same with 2 centroids in step 3, then the iteration is stopped.

Calculation of Value Sum Square Error (SSE)

In this phase will be calculated from the centroid value iteration results, if the results of the SSE value gets smaller, the better the results clusteringnya. There will be 2 testing parameters in calculating the value of SSE, the results will be used as a determinant of the cluster where the most good.

SSE

3 Centroid 461,004.5

2 Centroid 528,853.25

Based on the calculation of the smallest value of SSE SSE is the value of 3 centroid, it can be concluded as the best cluster in a research trial in the case of training data from 2008 to 2013 prediction of pneumonia disease that strikes children under five years in the future in Malang, due to the impact of frequency Exclusive breastfeeding, immunization status and nutrient intake.

G. Cluster Determination of Initial Training Data

Based on the calculation of the smallest value of SSE SSE is the value of 3 centroid, it can be concluded as the best cluster, so as to create two clusters based on the training data as follows:

Cluster Pertama (C0) : Rendah

TAHUN	ASI EKSKLUSIF	IMUNISASI	ASUPAN GIZI
2008	459	631	288
2009	356	489	539

Cluster Kedua (C1)

TAHUN	ASI EKSKLUSIF	IMUNISASI	ASUPAN GIZI
2010	493	187	629
2011	265	614	379
2012	297	244	573
2013	834	402	106

Figure 10. Cluster Final Determination

Based on the results of the K-Means clustering, then the pattern of the results obtained from each cluster. Here's an explanation of each cluster on each member of the cluster:

• Members of the data on the first cluster (C0) has a characteristic value for the problem at the estimated value of exclusive breastfeeding from 459 to 356, for the immunization of the estimated value of 631 to 489, and for the nutritional intake of 288 to 539.

• Members of the data on the second cluster (C1) has a characteristic value for the problem at the estimated value of exclusive breastfeeding from 439 to 834, for the immunization of the estimated value of 187 to 402, and for the nutritional intake of 629 to 106.

Clustering results of K-Means prediction can be illustrated in the following chart:

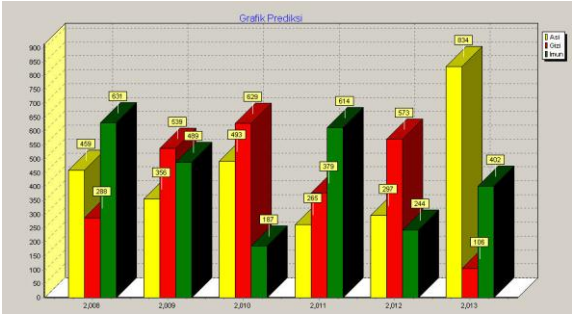

Figure 11. Prediction of Final Determination of Cluster Graphs

V. CONCLUSION

As for the conclusion that the bias obtained from research progress report K-Means Method To Learn The high Pneumenia Disease Based Attacks Toddler Time Range of exclusive breastfeeding, immunization status and Nutrition:

• There is a relationship between the time range of exclusive breastfeeding, immunization and nutrition status of the low height of the toddler pneumonia.

• With the K-Means method can predict the level of toddlers aged patients with pneumonia associated with variable time range of exclusive breastfeeding, immunization and nutrition status.

VI. SUGGESTION

Advice from the results of the progress of this study are as follows:

• Training data can be copied again to get more accurate results.

• Aspects supporting predictions can be developed even more with the development of pneumonia symptoms experienced by the community.

3 . Applications are made to be developed in the form of early detection of disease pneumonia using online media so that the public can easily access it.

REFERENCES

[1] ___. 2013. The benefits of exclusive breastfeeding for the baby. http:// http://www.botolkacaasi.com. March 11, 2013

[2] Fanada, Mery and Young, Widyaiswara. 2012. Factors Associated With Disease Occurrence Pneumonia In Toddlers In Palembang Kenten Work Area Health Center in 2012.

Provincial Training Agency. South Sumatra

[3] Nango, Dwi Noviati. 2012. The application of K-Means Algorithm for Clustering Data Local Budget In District Xyz. Department of Informatics. University of Gorontalo

[4] Rahman, Vionny. 2012. Do you know what are the contents of the milk?. http://www.duniaanak.org. March 11, 2013.

[5] Santoso, Budi. 2007. Applied Data Mining with MATLAB. First Edition. Graha Science. Yogyakarta

[6] Widyawati, N. 2010. Comparison of Clustering Based On frequest Word squence (CFNS) and K-Means clustering Documents To Speak Indonesian . Faculty of Mathematics and Natural Sciences. UPI. duo

[7] Yafonda, Dea. 2011. Nutritional status and breastfeeding in Genesis Against Pneumonia In Toddlers Work Area Health Center Ambacang Kuranji Padang District of the Year 2011. Faculty of Nursing. Universitas Andalas