

Classification of Low Birth Weight Baby Under Anthropometry uses Algorithms K-Means Clustering on Maternity Hospital

Irfan Santiko^{a,1,*}, Deni Kurniawan^{b,2}

^{a,b} Informatic Departement Universitas AMIKOM Purwokerto, Let.Jend. Pol Sumarto, Purwokerto, 53113, Indonesia

¹ irfan.santiko@amikompurwokerto.ac.id *; ² denikurnia0112@gmail.com

* corresponding author

Abstract

LBW infants with birth weight less than 2500 grams regardless gestation period. Low birth weight is the weight of a baby who weighed within 1 hour after birth. World Health Organization (WHO) since 1961 states that all newborns are underweight or equal to 2,500 g called low birth weight infant (low birth weight). According to WHO. Statistically, morbidity and mortality in neonates in developing countries is high, with the main causes is associated with LBW. To facilitate medical personnel in determining the risk of LBW. From the testing that has been done by the author, the k-means clustering algorithm has accuracy in classifying LBW babies by spacing the proximity between variables and the similarities in the test data,

Keywords: LBW; Clustering; K-means;

1. Introduction

The informatic technology in this current era is evolving This has implications for daily life that require data or information very quickly and accurately. Where the technology is needed dikehidupan human being in the present and even future. Due to the use of technology in the present very much in use, as an example in the field of health. Especially the health of the baby. In case this is a case of low birth weight (LBW) [8].

Weight loss is one of the indicators of the health of the Newborn (BBL). Average normal weight (getasi age 37 till 41 weeks) is 3200 grams. In general, low birth weight greater risk for having problems or complications at birth [8].

LBW infants with birth weight less than 2500 grams regardless getasi period. Low birth weight is the weight of a baby who weighed within 1 hour after birth. World Health Organization (WHO) since 1961 states that all newborns are underweight or equal to 2,500 g called low birth weight infant (low birth weight babies). According to WHO LBW babies are born weighing less than 2500 grams. Infant mortality rate is a key indicator in determining the health status of children, because it is a reflection of ststus child health at the moment. Statistically, morbidity and mortality in neonates in developing countries is high, with the main causes is associated with LBW [8].

LBW including major factor in increased mortality, morbidilitas, and disability neonatal infants, and children as well as provide long-term impact on his life in the future. LBW births continues to increase annually in developed countries like the United States, while in Indonesia it is followed by the birth of low birth weight infant deaths [8].

The birth rate of LBW in the world reached 14%. Developing countries occupy the birth rate of LBW by 15%, while the industrialized countries the birth rate of LBW 7%. Based on the results of the study Demographic and Health Survey (DHS) from 2002 to 2003 and reanalyzed by UNICEF HQ (Headquarter) in June 2004, pravalensi LBW births in Indonesia reached 9% [7].

Pravelensi LBW according to WHO (2010) estimated 15% of all births in the world with limits of 3.3% to 3.8% and is more common in developing countries or low income. Statistically showed 90% LBW obtained in developing countries with a mortality rate 35 times higher than in infants with a birth weight of more than 2500 grams. It can happen and can be affected by several factors such as the mother has the disease directly related to pregnancy, and age of the mother, [2].

The incidence in Indonesia vary considerably from one region to another, ie berkisan between 9% -30%, the study results obtained figures in 7 regions of LBW can be known based on estimates from the Survey of Demography and Health of Indonesia (IDHS). Nationally based on further analysis IDHS LBW figure of about 7.5%. This figure is greater than the target set out lbw on nutrition improvement program targeted towards Healthy Indonesia ie a maximum of 7%. [8].

While the anthropometric measurements plays pretty much based on the characteristic features of LBW infants, because anthropometry is measurement of human body dimensions and design applications involving physical

geometry, mass, strength and characteristics of the human body in the form of shapes and sizes. Humans will basically have the shape, height and weight are different from one another, for the research was conducted to classify LBW infants by anthropometric using k-means algorithm.

K-Means Clustering is a method of analyzing data or methods that perform data mining modeling process without supervision (unsupervised) and is one method of grouping the data to the system partition.

The conclusions analysis who obtained the best algorithm to perform clustering in the above studies is the K-Means algorithm, for grouping clusters in K-means algorithm is better than the EM algorithm [4].

The calculation results show K-Means method produces better results than using DBSCAN. Based on these studies, the authors chose to use an algorithm K-Means clustering method in this study to classify low birth weight infants. So it can be found certain patterns that yield important information that support the process of checking LBW [1].

On this topic based on the above background can be taken the problem in this research is, how the results grouping LBW infants by anthropometric measurements using data mining with k-means clustering algorithm. The aim of this research is the grouping infants included in the diagnosis of LBW to help midwives and medicine in determining LBW infants groups based on the calculation of k-means clustering algorithm.

2. The Proposed Method and Literature

2.1. Data Mining

Data mining is a series of processes for adding additional value of a set of data in the form of knowledge that had been unknown to them manually [3].

2.2. K-Means Algorithms

K-Means clustering is a method of group analysis that led to the object of observation of partitioning N into K groups (clusters) where each object observation is owned by a group with a mean (average) nearby [5]. the steps do clustering with K-Means method as follows :

- a. Select the number of clusters k.
- b. Initialize k cluster center with random manner.
- c. Data to calculate the distance to each centroid using distance formula Euclidean Distance. Here is the equation of Euclidean Distance:

$D(i, j)$

$$= \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Where:

$D(i, j)$ = Distance Data to i to the cluster center j

X_{ij} = Data to i the data attributes to k

X_{kj} = Data to j the data attributes to k

- d. Recalculate center of the cluster with the current cluster membership.

Determine each object again wear the new cluster center. If the cluster centers do not change again the clustering process is completed. Or, go back to step c until the cluster centers do not change anymore.

2.3. Anthropometry

Anthropometry is derived from the "anthro" means human and "metron" meaning measure. Anthropometric definitively declared as a study concerning the measurement of human body dimensions and design applications involving physical geometry, mass, strength and characteristics of the human body in the form of shapes and sizes. Humans will basically have the shape, height and weight are different from one another [9].

2.4. Low Birth Weight (LBW)

Birth weight is the weight of neonates at birth weighed within one hour after the body lahir. berat anthropometric measure the most important and most frequently used in newborns (neonates). Weight loss is used to diagnose normal or low birth weight infants [2].

2.5. WEKA Tools

Weka is an open source data mining applications based on java. This application was first developed by the University in New Zealand named the University of Waikato in 1994 [7].

3. Method

In this research using primary data, ie data Antropometri birth of Maternity Hospital Indonesian. Here is the flow of research using k-means clustering algorithm. This step for fulfillment :

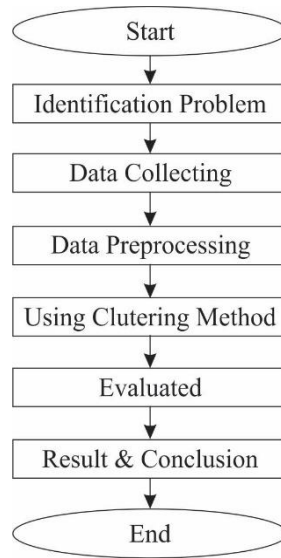


Figure 1. Concept Research

From the picture above, the lines of inquiry that are used as follows:

3.1. Identification Problem

The process of problem identification is performed to determine the problems and methods that would allow them to ditentuka points to classify infants into the category of LBW.

3.2. Data Collecting

In this research uses data Anthropometric birth of Maternity Hospital on Purwokerto Indonesian teritorial.

3.3. Data Preprocessing

At this stage, a data transformation process that is changing the nominal data into numerical data.

3.4. Using Clustering Method

To the use of clustering algorithms grouping Kmeans to get results based on data Anthropometric LBW infants.

3.5. Evaluated

In this phase all evaluated from the beginning to get the grouping which is processed using k-means clustering algorithm.

3.6. Result and Conclusion

The next stage is to conclude the results obtained from research based algorithm K-Means clustering that delivers results instagram social media use and value of students.

4. Results and Discussion

The sample used in this research were 1210 babies of 2016. Table 1 shows the data babies sample data.

Table 1. Adjustment Result Dataset

No.	Weight	Height	LK	LD	LILA
1	3490	48	33	32	11
2	3580	48	36	31	10
3	3320	48	33	32	11
4	3390	47	33	32	10
5	3590	47	34	32	11
6	3420	48	33	32	11
7	3410	49	35	32	11
8	3460	50	33	32	11

1208	3090	46	33	34	11
1209	2850	45	33	31	9
1210	3230	47	32	31	10

(Source: Data on Maternity Hospital Purwokerto)

4.1. Data Preparation

At the stage of data preparation is to convert the data LBW from a CSV (Comma-Separated Values) or Microsoft Excel into a file with the extension ARFF that can be easily opened by software weka, data samples LBW in CSV and data conversion can be seen in the picture below.

```

3490,48,33,32,11
3580,48,36,31,10
3320,48,33,32,11
3390,47,33,32,10
3590,47,34,32,11
3420,48,33,32,11
3410,49,35,32,11
3460,50,33,32,11
3870,50,35,33,10
Etc .....
    
```

Figure 2. Preparation of data

4.2. Handling Missing Values

At the stage of preprocessing the data, the next step needs to be done after the data set is to perform the data processing to handle the data subject to a missing values. In the data LBW data are missing values experienced in column 640 are attribute to weight attributes and attribute column to which the attribute LK 985 (Round Head). Below is a table of 2 missing data values.

Table 2. missing values

Weight	long Firm	LK	LD	Lila
3360	?	34	33	11
3320	43	?	27	9

(Source: Data processed LBW)

Below is described the calculation using the average method for handling missing values. The results of calculation of the value of substitution on record missing values as follows:

$$\begin{aligned}
 \text{Mean} &= \frac{\sum \text{nilaiattribute Panjang badan}}{1210} \\
 \text{Mean} &= \frac{48 + 47 + 49 + 50 + \dots + 46 + 47}{1210} \\
 \text{Mean panjang badan} &= 47 \\
 \text{Mean} &= \frac{\sum \text{nilaiattribute ke LK}}{1210} \\
 \text{Mean} &= \frac{33 + 36 + 32 + 34 + \dots + 33 + 32}{1210} \\
 \text{Mean LK} &= 33
 \end{aligned}$$

In this formula the results of calculations to find the mean value of the attributes of the body length LK data missing values experienced in body length and LK attribute is filled with the average value of all these attributes. Based on the average value obtained grades 47 and 32. Thus, a body length attribute filled with the value of 47 while LK attribute filled with the value of 32.

Table 3. The data of handling missing values

NO	Weight	long Firm	LK	LD	Lila
1	3360	47	34	33	11
2	3320	43	33	27	9

(Source: the data that has been processed LBW)

4.3. The use of clustering methods

After going through the preprocessing phase completed, the dataset start calculation and processed using weka apilkasi. The use of the above method is just to do grouping, the grouping obtained from the calculation iterations that will be performed by researchers. The data used in this study as many as 1210 data. At this stage, the author uses primary data to pengolaha data, using anthropometric data childbirth.

```

Run information === ===
Scheme: weka.clusterers.SimpleKMeans -N 5 -A "-R
weka.core.EuclideanDistance first-last" -I 500 -S 10
Relation: a data-weka.filters.unsupervised.attribute.Remove-R1
Instances: 1210
Attributes: 5
Weight
panjang_badan
lingkar_kepala
chest size
upper arm circumference
Test mode: Evaluate training on the data
=== Model and evaluation on the training set ===
Kmeans
=====
Number of iterations: 28
Within cluster sum of squared errors: 19.66543386890737
Globally replaced missing values with mean / mode
Cluster centroids:
cluster #
Attribute Full Data 0 1 2 3 4
(1210) (249) (328) (104) (323) (206)
=====
=====
Berat_badan 3126.9091 2967.8715 2939.0854 2408.5577 3346.9969
3635.7767
panjang_badan 47.2525 46.9036 46.1692 43.8173 48,678 48.8981
lingkar_kepala 33.1459 31.5803 33.7744 30.7885 33.1486 35.2233
lingkar_dada 31.0508 30.3795 30.6387 27.7885 31.8901 32.8495
lingkar_lengan_atas 10.3132 9.9819 10.0183 9.2404 10.7755 11

Time taken to build the model (full training data): 0.1 seconds
=== Model and evaluation on the training set ===
clustered Instances
0249 (21%)
1328 (27%)
2104 (9%)
3323 (27%)
4206 (17%)
    
```

Figure 4. Calculation Results anthropometric data, (Source: Data processed LBW)

Following testing grouping calculation based k-means method using weka application. The following chart readings weka.

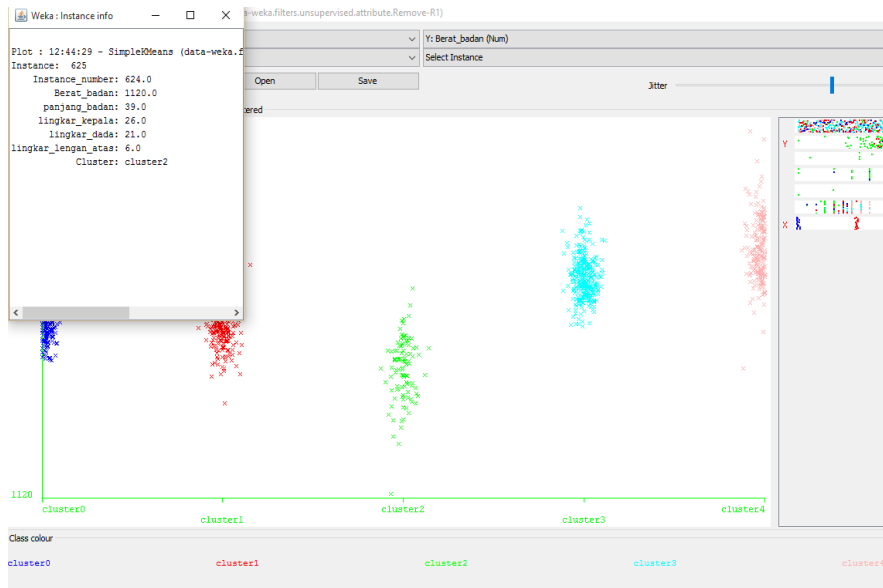


Figure 5. Graph weka

Based on Figure 4.2 and Figure 4.3 can be seen groups of infants who indicated LBW on cluster-2, 104 infants with a percentage of 9%, and cluster into two levels of potential LBW highest compared to cluster the other, because the data group of babies inside cluster 2 has the physical characteristics of the most low or below normal limits compared to all the data clusters baby tested. Based on the results of the cluster above, we can see that the cluster data to a second value lower than the anthropometry of infants most other clusters, for the second cluster is a group of LBW infants, following anthropometric data of cluster 2:

Table 4, Sample data from cluster 2

No.	Weight	body length	head circumference	Chest size	LILA	cluster
624	1120	39	26	21	6	cluster2
337	1550	39	26	24	8	cluster2
748	1830	43	29	25	7	cluster2
361	1890	39	32	26	8	cluster2
484	1920	44	31	28	9	cluster2
29	1930	42	31	24	8	cluster2
83	1940	43	30	25	8	cluster2
202	1950	40	29	27	9	cluster2
132	1960	44	33	27	7	cluster2
670	1980	42	30	25	7	cluster2
816	2080	46	32	26	8	cluster2
336	2130	40	30	27	9	cluster2
48	2140	43	30	26	8	cluster2
153	2140	43	32	28	9	cluster2
566	2140	40	30	26	5	cluster2
1067	2150	43	30	27	9	cluster2
218	2190	44	32	28	9	cluster2
444	2190	44	32	29	9	cluster2
532	2190	47	30	28	08:05	cluster2
257	2200	44	30	28	9	cluster2

.....
.....
220	2990	43	31	28	9	cluster2
514	3000	45	30	27	9	cluster2
84	3230	45	31	5	31	cluster2

(Source: Data LBW already in clusterkan)

5. Conclusion

In this study, we concluded grouping of testing k-means clustering algorithm. From the testing that has been done by the author, then the algorithm k-means clustering to have accuracy in classifying LBW infants berdasarkan distance of proximity between variables and the similarities in the test data, such as results already in the know, that the baby is diagnosed LBW everything into the cluster group second, that there are 107 data or 9% of LBW dataset in which there are indications of LBW babies, while another group is a group of babies are normal and do not get into the group of LBW.

Acknowledgment

From the analysis conducted by the authors, to categorize LBW infants, the authors provide suggestions as follows, further testing can be maximized by adding a data sample that has a variable or attribute that is more complete and expected to be implemented into a system.

References

- [1] Budiman, et al. 2016. Comparison of Methods K-Means and Methods
- [2] DBSCAN At Boarding House Grouping Students in the Village Tembalang Semarang. *Journal Gaussian*. Vol. 5 No. 4 Year 2016 Page 757-762. ISSN: 2339-2541.
- [3] Kusriani and Lutfhi, ET 2009. *Data Mining Algorithm*. Andi Yogyakarta.
- [4] Mardiani. 2014. Comparison Algorithm K-Means and EM for Clusterisasi Value Based Home School Students. *CITEC Journal*. Vol. 1 No. 4 August to October 2014. ISSN: 2354-5771.
- [5] Prasetyo, Eko. 2012. *Data Mining - Concepts and Applications Using MATLAB*. ANDI OFFSET.
- [6] Pramudyo Eko Rio. "Sistem Pakar Diagnosa Penyakit Infeksi Saluran Pernafasan Akut (ISPA) Dengan Menggunakan Metode Sugeno dan Naive Bayes".
- [7] Mastian, Hasni. "Faktor-Faktor yang Berhubungan dengan Kejadian Berat Badan Lahir Rendah." *NERS Jurnal Keperawatan* 7.2 (2011): 114-120.
- [8] Alya, dian., And STIK U'BUDIYAH. "Factors associated with low birth weight (LBW) in maternal and child hospital Banda Aceh in 2013." Essay. Midwifery Diploma Course IV STIKes Ubudiyah: Banda Aceh (2014).
- [9] Swastina, Liliana, and Bambang Lareno. "Model Penentuan Potensi Status Gizi Bermasalah Menggunakan Decision Tree." (2014).
- [10] Patil, Tina R., and S. S. Sherekar. "Performance analysis of Naive Bayes and J48 classification algorithm for data classification." *International Journal of Computer Science and Applications* 6.2 (2013): 256-261.
- [11] C. Boutsidis, A. Zouzias, P. Drineas, Random projections for k-means clustering, in: *NIPS*, 2010, pp. 298-306.
- [12] P.S. Bradley, U.M. Fayyad, Refining initial points for k-means clustering, *ICML 98* (1998) 91-99.
- [13] M.B. Cohen, S. Elder, C. Musco, C. Musco, M. Persu, Dimensionality reduction for k-means clustering and low rank approximation, in: *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, 2015, pp. 163-172.
- [14] I. Davidson, A. Satyanarayana, Speeding up k-means clustering by bootstrap averaging, *IEEE data mining workshop on clustering large data sets*, 2003.
- [15] V. Gandhi, J.M. Kang, S. Shekhar, *Spatial Databases*, Wiley Encyclopedia of Computer Science and Engineering, 2008.