

CASE BASED REASONING (CBR) FOR OBESITY LEVEL ESTIMATION USING K-MEANS INDEXING METHOD

^aI Made Satria Bimantara, ^bI Wayan Supriana

^{a, b}Informatics Department, Faculty of Math and Science, Udayana University, Bukit Jimbaran, Bali, Indonesia-803611, Telp: 081237787141
E-mail: satriabimantara.md@gmail.com, wayan.supriana@unud.ac.id

Abstract

As many as 600 million of the 1.9 billion adults who are overweight are obese. Obesity that is not treated immediately will be a risk factor for increasing cardiovascular, metabolic, degenerative diseases, and even death at a young age. Case Based Reasoning (CBR) can be used to estimate a person's obesity level using previous cases. The old case with the highest similarity will be the solution for the new case. Indexing methods such as the K-Means Algorithm are needed so that the search for similar cases does not involve all cases on a case base so that it can shorten the computation time at the retrieve stage and still produce optimal solutions. Cosine similarity is used to find relevant clusters of new cases and Euclidean distance similarity is used to calculate similarity between cases. Random subsampling method was used to validate the CBR system. The test results with $K=2$ indicate that the CBR is better than the CBR-K-Means, each of which produces an average accuracy of 88.365% and 88.270% at a threshold of 0.8. CBR-K-Means produces an average computation time at the retrieve stage of 33.55 seconds and is faster than the CBR of 35.5 seconds.

Key words: Case Based Reasoning, Case Indexing Method, K-Means, Obesity.

INTRODUCTION

Obesity is an excessive accumulation of fat due to an imbalance of energy intake with energy used for a long time [1]. Ng added that obesity is thought to occur due to lack of physical activity, genetic factors, stress and others. As many as 62% of obese people in the world are in developing countries. The state of Indonesia is one of them [2]. Other factors such as drugs, environment, and hormones can be the cause of obesity [1].

The Ministry of Health of the Republic of Indonesia (Kemkes RI) revealed the fact that obese people in the world have more than doubled since 1980. More than 1.9 billion adults aged 18 years and over are overweight.

More than 600 million are obese of this number [3]. Based on the results of the Riset Kesehatan Dasar (Riskesdas), there was an increase in the prevalence of obesity in the population aged over 18 years from 11.7% in 2010 to 15.4% in 2013 in Indonesia [1]. Based on the 2015-2019 RPJMN indicators, 15.4% are obese with a BMI 27 in Indonesia. In adults aged 18 years and over, 13.5% were overweight. Meanwhile, 28.7% were obese with a BMI 25. In children aged 5-12 years, 18.8% were overweight and 10.8% were obese [3].

According to WHO there are five levels of obesity based on the measurement of Body Mass Index (BMI). The levels are

underweight, normal weight, overweight with risk, obesity I, and obesity II [4]. In addition to using BMI measurements, obesity levels can be determined from dietary habits and physical condition. Levels of obesity are divided into seven levels, namely underweight, normal weight, overweight level I, overweight level II, obesity level I, obesity level II, and obesity level III [5].

Obesity cannot be underestimated. Hubby et al. [2] explains, there are health impacts caused by obesity in adults. The increase in the incidence of non-communicable diseases such as type 2 diabetes, cancer, other cardiovascular diseases, and even death at a young age is caused by obesity as a risk factor. Kemkes RI added that obesity that is not immediately treated will be a risk factor for metabolic diseases (triglycerides, decreased HDL cholesterol, and increased blood pressure), degenerative diseases, and other diseases (exacerbation of asthma, knee and hip osteoarthritis, gallstone formation, cardiac arrest). breath during sleep, and low back pain) [1], [3]. Therefore, it is important to know a person's obesity level early so that appropriate treatment can be carried out by experts. Case Based Reasoning (CBR) system can be used to estimate a person's obesity level based on eating habits and physical activity using previous cases.

CBR is a method used to study and solve a problem based on past experiences. This method uses a case-based approach [6]–[8]. The solution of a problem can be determined by finding a solution of the old case in a similar case base. If a case in the case base is similar to an old case, it can be said that both are identical [9]. CBR has been widely implemented in several ways, such as cases of caesarean section [8], diagnosis of hypertension [10], prediction of length of study for prospective students [11], diagnosis of heart disease [9], diagnosis of stroke [6], [12], and determine deviant sexual behavior [13]. CBR will be used to estimate the level of obesity based on dietary habits and physical condition in this study.

The process of finding old cases that are similar to new cases is a problem with CBR. The computation time will be longer when the system has to calculate the similarity value of the new case to all the old cases in the case base. Indexing of old cases is needed to handle this [10]. The process of calculating the similarity value will not involve all of the old cases in the case base, but only the old cases closest to the new case. With the indexing of old cases is expected to shorten the computation time and still produce optimal solutions.

The proposed of indexing method in this study uses a partition clustering algorithm, namely K-Means Algorithm. With the K-Means Algorithm, several data will be grouped based on similarities and dissimilarities into groups called clusters. Thus, each cluster will contain a set of data with a high level of similarity and a low level of dissimilarity. The implementation of the algorithm is expected to increase the efficiency of computing time, especially during the retrieval stage and still produce an optimal solution.

MATERIAL AND METHODS

Flowchart System

The flowchart of the CBR system is shown in Figure 1.

Research Data Collection

This study uses secondary data obtained from the UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets/Estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition+>) with the name "Estimation of Obesity Levels Based On Eating Habits and Physical Condition Data Set". This data consists of 2,111 records and 17 attributes. This data is estimation of obesity levels in individuals from Mexico, Peru and Colombia based on their dietary habits and physical condition. The data has been classified into seven levels of obesity [5].

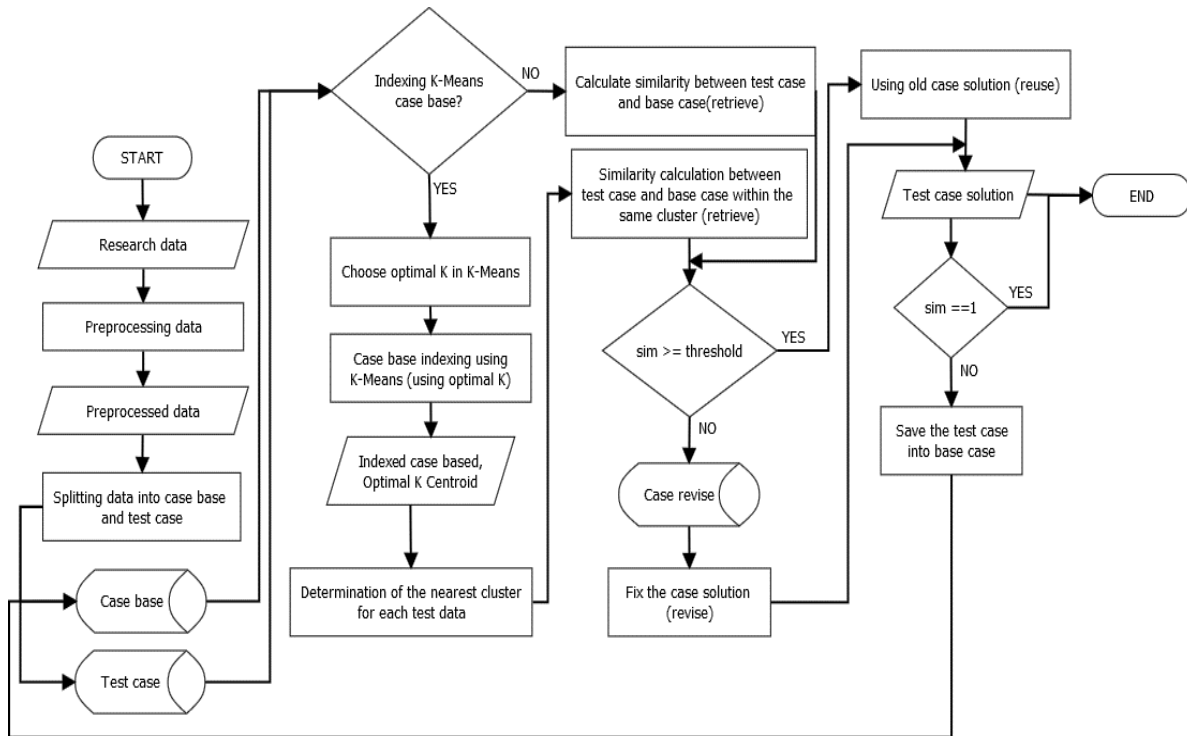


Fig 1. Flowchart of the case based reasoning system

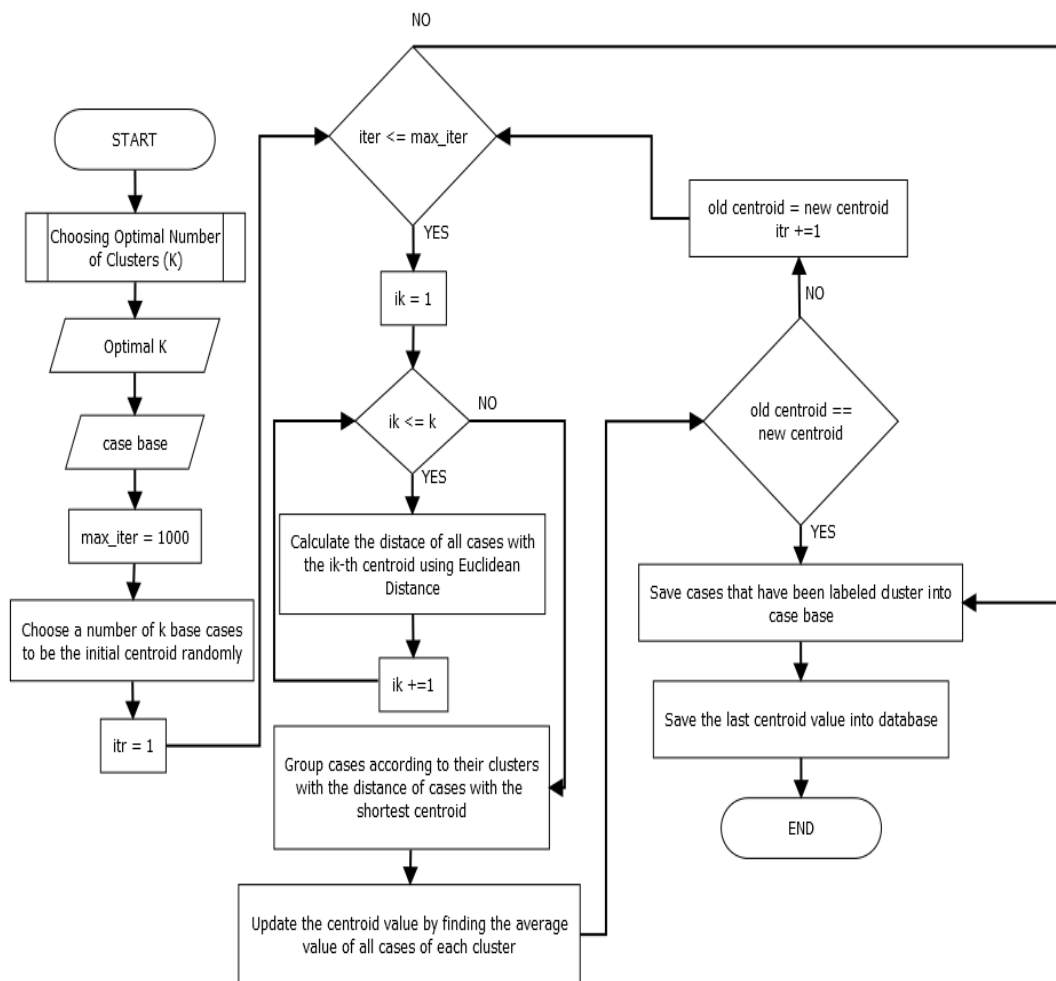


Fig 2. Flowchart of case base indexing process using K-Means clustering algorithm

Preprocessing Data

Data Cleaning

Data cleaning is carried out on data that is still dirty. Data is said to be dirty if the data has an empty value (*missing value*) and/or outliers and/or inconsistent [14]. Missing values and outliers will be handled by discarding the data and not being used in subsequent data processing. There are no missing values and outliers after checking the data.

Label Encoding

Label Encoding is a method for converting data with categorical data types (nominal and ordinal) into numeric data types. Most of the attributes in the data have a category data type, therefore this process is carried out. The purpose of this process is so that each stored case can later be indexed using the K-Means Algorithm. This process is only carried out for attributes of type category except for attributes that are the target class of the data.

Data Normalization

Data normalization is a process to change the attribute values in the data into the same range. Normalization is usually done into the range [0,1] or [-1,1], so that each attribute will have the same weight and avoid bias [15]. The normalization strategy used in this study uses min-max normalization according to Equation (1). This strategy was chosen because it provides the highest average accuracy value compared to other normalization methods such as z-score normalization and decimal scaling [16], [17]; and sigmoid, softmax, and statistical column [16].

$$y = \frac{x - D_{min}}{D_{max} - D_{min}} \quad (1)$$

Determination of the Optimal Number of Clusters

Determination of the optimal number of clusters is carried out before the case base is indexed using K-Means. The Elbow method will be used at this stage. The Elbow method determines the optimal number of clusters by getting the largest difference value from the decrease in Sum of Square Error (SSE) and forming an angle. SSE is the sum of squared errors between all data in cluster C_j and the cluster center point (centroid) P_j . SSE is calculated using Equation (2).

$$SSE = \sum_{j=1}^K \sum_{x_i \in C_j} \|x_i - P_j\|^2 \quad (2)$$

Case Representation

The case representation model used is flat frame. The attributes and solutions of the cases will be represented as in Table 1. There are 17 attributes and 7 solution spaces (target classes). A row named *Cluster* added to the case representation to store the relevant cluster values of the case.

Base Case Indexing Using K-Means Algorithm

The K-Means Algorithm is used as a method for indexing old cases in the case base. The number of k optimal groups formed was obtained using the Elbow method. After the number of k optimal is obtained, all the old cases on the case base will be indexed according to the cluster of cases. The case base indexing flow shown in Figure 2. The new case will look for the similarity value from the old cases that are in the same cluster as it.

New Case's Relevant Cluster

CBR System with K-Means indexing will determine the closest cluster to the new case to be able to find a solution. The closest cluster is obtained by calculating the similarity of the new case to the cluster center point (centroid) that has been generated in the previous stage. The new case and the centroid are represented as vectors and the similarity is calculated using the equation cosine similarity shown in Equation (3). The two vectors (cases) are said to be similar if the result of cosine value is equal to one, otherwise if the value is zero, then the two vectors are said to be dissimilar [10].

$$\begin{aligned} \text{cosine}(\bar{u}, \bar{v}) &= \frac{\bar{u} \cdot \bar{v}}{\|\bar{u}\| \|\bar{v}\|} \\ &= \frac{\sum_{i=1}^N (\bar{u}_i \times \bar{v}_i)}{\sqrt{\sum_{i=1}^N \bar{u}_i^2} \times \sqrt{\sum_{i=1}^N \bar{v}_i^2}} \end{aligned} \quad (3)$$

Descriptions:

$\text{cosine}(\bar{u}, \bar{v})$: *cosinus* value of new case \bar{u} dan *centroid* \bar{v}

\bar{u} : new case vector

\bar{v} : *centroid* vector

N : number of attributes in vector

\bar{u}_i : value of i -th element in vector \bar{u}

\bar{v}_i : value of i -th element in vector \bar{v}

Case Similarity

Euclidean distance similarity method used to calculate similarity during the retrieval stage. This method measures the similarity of the new case to the old case by calculating the closeness of the distance between the two case based on the similarity of weights on each of the existing attributes [10], [11], [13]. Similarity measurements are carried out, namely local similarity and global similarity. Local similarity is a measurement of similarity at the attribute level, while global similarity is a measurement of similarity at the level of the whole case [9], [10], [12], [18].

Local Similarity Measurement

Local similarity measurement in this study is distinguished for numeric and ordinal type attributes; as well as nominal and binary. Local similarity for numeric and ordinal type attributes is calculated using Equation (4), while for nominal and binary type attributes is calculated using Equation (5) [9], [10], [12].

$$f(S_i, T_i) = 1 - \frac{|S_i, T_i|}{|f_{max} - f_{min}|} \quad (4)$$

$$f(S_i, T_i) = \begin{cases} 1, & S_i = T_i \\ 0, & S_i \neq T_i \end{cases} \quad (5)$$

Descriptions:

$f(S_i, T_i)$: local similarity of i -th attribute between case S with case T

S_i : i -th attribute value of case S

T_i : i -th attribute value of case T

f_{max} : maximum value of i -th attribute value in case base

f_{min} : minimum value of i -th attribute value in case base

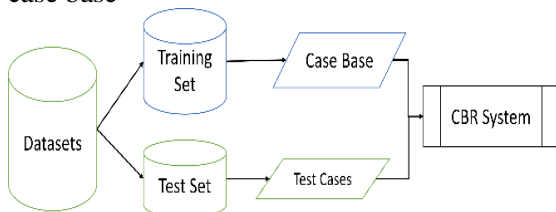


Fig 3. Illustration of holdout method for dividing data into cases base and test cases

Global Similarity Measurement

Measurement of global similarity between new cases and old cases that are in the case base using Equation (6) [9], [10] with calculation *Euclidean distance*. The weight for each attribute uses *Pearson Correlation* to the target class obtained through correlation analysis by ignoring the positive and negative signs.

Calculation of correlation analysis of each attribute to the target class using IBM SPSS Statistics 26 software. The calculation results are shown in Table 2. Att represents the list of attributes used. Pc represents the Pearson correlation value. Scp represents the sum of squares and cross-products. Cov represents the covariance value.

$$sim(S, T) = \left(\frac{\sum_{i=1}^N w_i^2 \times |f(S_i, T_i)|^2}{\sum_{i=1}^N w_i^2} \right)^{1/2} \quad (6)$$

Descriptions:

$sim(S, T)$: similarity new case T with source case S

w_i : weight of i -th attribute

$f(S_i, T_i)$: local similarity of i -th attribute between case S and case T

N : attribute number in the vector

Evaluation and Validation of Results

The output results of the CBR system will be evaluated using an accuracy measure. The accuracy generated by the CBR system for estimating the level of obesity will be determined based on the percentage of the number of cases estimated correctly according to Equation (7). This accuracy value will be used to compare the performance of the CBR system using K-Means indexing method and without indexing method.

$$acc = \frac{\sum \text{number of correct cases}}{\sum \text{number of test case}} \times 100\% \quad (7)$$

Holdout method used to divide research data into case base and test case. The data set D is divided randomly into two independent subsets of data, namely training data and test data [14], [15], [19]. The training data will be used as the case base, while the test data will be used as test cases to compare the performance of the CBR system with and without K-Means indexing method. This study uses the proportion of training data of 80% and test data of 20% [20]. The illustration of Holdout method used in this study is shown in Figure 3.

Validation of the results of the CBR system will use random subsampling method. This method is a Holdout method that run some of k iterations. The average value of accuracy to k iterations will be used to compare the performance of CBR systems with and K-Means indexing method. In this study the number of iterations was set at 20 iterations.

Analysis of Data Processing Results

Analysis of data processing results is carried out at the final stage of this research. The thing that is analyzed in this study is to compare the computation time and accuracy values generated by the CBR system with and without K-Means indexing method.

RESULT AND DISCUSSION

Two test scenarios were used to analyze the ability of the CBR system to estimate obesity levels. Each scenario will be carried out 20 iterations. The first scenario is system testing using CBR without K-Means indexing method and the second scenario is system testing using CBR with K-Means indexing method. Cosine similarity method is used to find clusters that are relevant to the new case. Euclidean distance is used to determine the similarity of the new case to the case base.

System Implementation

Figure 4 shows the implementation of the CBR system that has been made based on the website. The figure presents the results of the comparison of CBR with and without K-Means indexing method. Figure 5 shows the solution of the CBR system with and without K-Means indexing method for each tested test case.

Testing the optimal number of clusters

The Elbow method is used to test number of clusters or k . The test was carried out 10 times. The highest average SSE difference from 10 times of testing will be the optimal value of k . This test will use a range of values k from 1 to 10 [19]. The test results for the optimal number of clusters are shown in Table 3 and Table 4.

Based on the results in Table 4, the value k equal to two gives the highest result for the average value of the SSE difference, which is 2631.27 with a standard deviation of 6.95. The

value of k equal to two is the point that forms the angle in Figure 6. Therefore, the number of optimal k used in the CBR test with and without K-Means indexing method is two.

Testing the Comparison of CBR and CBR-K-Means

This test was conducted to see a comparison of the performance of the resulting CBR system. The value of accuracy (in percent) and computation time at the retrieval stage (in seconds) will be compared between the CBR system with and without K-Means indexing method. The results of this test are shown in Table 5 and illustrated in graphical form in Figure 7 for the accuracy value and Figure 8 for the computational time at the retrieval stage.

Based on Table 5, the CBR system without K-Means indexing method provides an average accuracy of 88.365% and is better than the CBR system with K-Means indexing method of 88.270%. The average computation time at the retrieval stage of the CBR system with K-Means indexing method is 33.55 seconds, faster than the CBR system without K-Means indexing method of 35.5 seconds.

In the CBR system with K-Means indexing method, the search for relevant cases only occurs on the case base that have the same cluster as new cases. Therefore, the computation time at retrieval stage becomes faster. The accuracy value generated by the CBR system has not provided 100% optimal results because the system has not been able to handle two or more cases that have the same similarity value as the new case. If this happens, the CBR system will take the case in the first serial number to be used as a new case solution which will have implications for the resulting accuracy value.

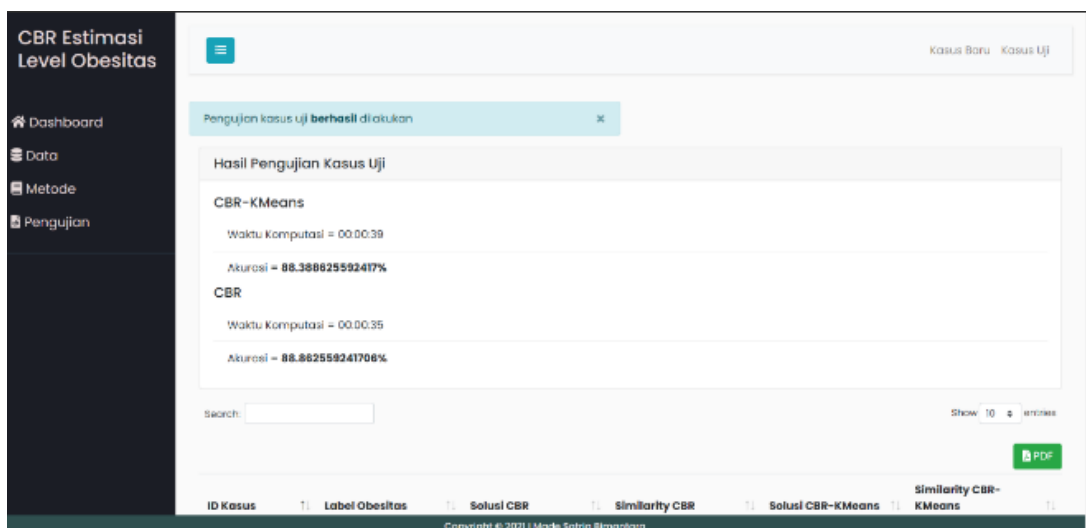


Fig 4. Implementation of the website based CBR system

| ID Kasus | Label Obesitas | Solusi CBR | Similarity CBR | Solusi CBR-KMeans | Similarity CBR-KMeans |
|----------|---------------------|---------------------|----------------|---------------------|-----------------------|
| 1 | Normal Weight | Overweight Level II | 0.997412 | Overweight Level II | 0.997412 |
| 5 | Overweight Level II | Overweight Level II | 0.989641 | Overweight Level II | 0.989641 |
| 14 | Obesity Type I | Obesity Type I | 0.95823 | Obesity Type I | 0.95823 |
| 17 | Overweight Level II | Overweight Level II | 0.993525 | Overweight Level II | 0.993525 |
| 23 | Normal Weight | Insufficient Weight | 0.990598 | Insufficient Weight | 0.990598 |
| 26 | Normal Weight | Normal Weight | 0.976513 | Normal Weight | 0.976513 |
| 28 | Normal Weight | Normal Weight | 0.9868 | Normal Weight | 0.97544 |
| 29 | Normal Weight | Overweight Level I | 0.980152 | Overweight Level I | 0.980152 |
| 35 | Normal Weight | Normal Weight | 0.983772 | Normal Weight | 0.983772 |
| 37 | Normal Weight | Insufficient Weight | 0.973786 | Insufficient Weight | 0.973786 |

Fig 5. The results of the solutions produced by the CBR system with and without K-Means indexing method

Table 3. The Sum of Squares Error (SSE) value in each i-th iteration for each cluster

| K | The SSE Value of the i-th Test | | | | | | | | | |
|----|--------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 4647,98 | 4647,98 | 4647,98 | 4647,98 | 4647,98 | 4647,98 | 4636,79 | 4636,79 | 4636,79 | 4636,79 |
| 2 | 2011,33 | 2011,33 | 2011,33 | 2011,33 | 2011,33 | 2011,33 | 2013,60 | 2013,60 | 2013,60 | 2013,60 |
| 3 | 1759,24 | 1653,07 | 1931,33 | 1759,24 | 1910,64 | 1989,90 | 1651,40 | 1777,71 | 1651,40 | 1651,40 |
| 4 | 1472,54 | 1490,39 | 1490,39 | 1651,08 | 1527,64 | 1490,56 | 1606,66 | 1607,11 | 1635,04 | 1496,95 |
| 5 | 1670,83 | 1399,55 | 1477,45 | 1387,60 | 1465,17 | 1387,60 | 1485,75 | 1532,90 | 1430,23 | 1395,69 |
| 6 | 1354,72 | 1382,52 | 1329,22 | 1384,13 | 1385,00 | 1328,55 | 1331,27 | 1359,34 | 1349,06 | 1423,50 |
| 7 | 1317,25 | 1355,22 | 1299,97 | 1305,57 | 1321,35 | 1331,74 | 1247,41 | 1246,31 | 1347,72 | 1410,10 |
| 8 | 1288,37 | 1410,87 | 1342,79 | 1293,13 | 1272,13 | 1328,84 | 1296,52 | 1274,37 | 1283,79 | 1231,36 |
| 9 | 1291,04 | 1178,32 | 1259,03 | 1250,19 | 1194,92 | 1246,72 | 1189,13 | 1225,95 | 1250,70 | 1153,66 |
| 10 | 1148,55 | 1167,66 | 1189,93 | 1191,62 | 1168,30 | 1114,89 | 1134,94 | 1185,95 | 1164,03 | 1216,76 |

Table 4. The Sum of Squares Error (SSE) difference value in each i-th iteration for each cluster

| K | The SSE Value of the i-th Test | | | | | | | | | |
|----|--------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 2 | 2636,65 | 2636,65 | 2636,65 | 2636,65 | 2636,65 | 2636,65 | 2623,19 | 2623,19 | 2623,19 | 2623,19 |
| 3 | 252,09 | 358,26 | 80,00 | 252,09 | 100,69 | 21,42 | 362,20 | 235,89 | 362,20 | 362,20 |
| 4 | 286,70 | 162,68 | 440,94 | 108,16 | 383,00 | 499,34 | 44,74 | 170,61 | 16,36 | 154,45 |
| 5 | 198,29 | 90,84 | 12,94 | 263,48 | 62,46 | 102,96 | 120,91 | 74,21 | 204,81 | 101,26 |
| 6 | 316,10 | 17,03 | 148,23 | 3,47 | 80,18 | 59,05 | 154,48 | 173,56 | 81,17 | 27,81 |
| 7 | 37,48 | 27,31 | 29,24 | 78,56 | 63,64 | 3,19 | 83,86 | 113,03 | 1,34 | 13,40 |
| 8 | 28,88 | 55,65 | 42,82 | 12,44 | 49,22 | 2,90 | 49,11 | 28,06 | 63,93 | 178,74 |
| 9 | 2,67 | 232,55 | 83,77 | 42,94 | 77,22 | 82,12 | 107,39 | 48,42 | 33,09 | 77,69 |
| 10 | 142,49 | 10,66 | 69,10 | 58,57 | 26,62 | 131,83 | 54,19 | 40,00 | 86,67 | 63,10 |

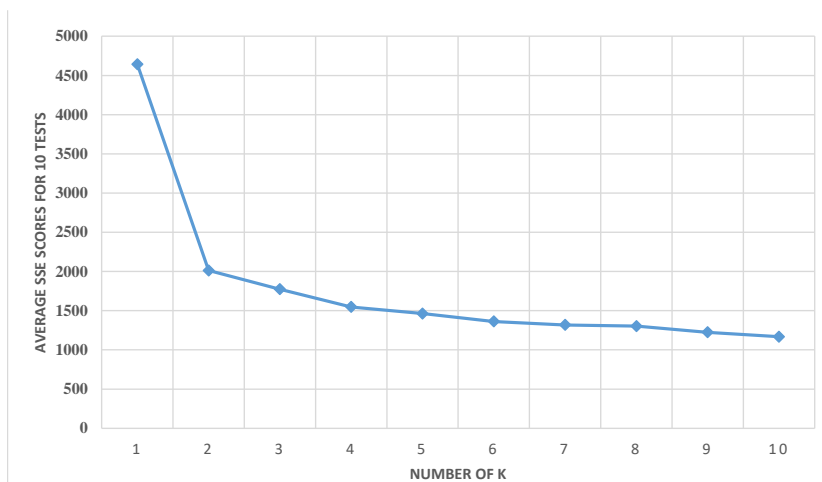


Fig 6. Graph of the average SSE value for every k for 10 times of testing

Table 5. Results of testing the CBR system with and without K-Means indexing method to estimate obesity level performed 20 times

| Number of Test | CBR-K-Means | | CBR | |
|-----------------|--------------|-------------------------|--------------|-------------------------|
| | Accuracy (%) | Computing time (second) | Accuracy (%) | Computing time (second) |
| 1. | 85,07109005 | 35 | 85,450237 | 35 |
| 2. | 88,62559242 | 31 | 88,15165 | 30 |
| 3. | 89,33649289 | 35 | 88,62559242 | 33 |
| 4. | 88,15165877 | 32 | 88,8625592 | 31 |
| 5. | 87,44075829 | 34 | 87,203791 | 35 |
| 6. | 89,57345972 | 35 | 89,81042654 | 37 |
| 7. | 88,15165877 | 34 | 87,91469194 | 36 |
| 8. | 84,5971564 | 33 | 85,07109005 | 35 |
| 9. | 90,04739336 | 34 | 89,33649289 | 36 |
| 10. | 86,72985782 | 33 | 88,15165877 | 36 |
| 11. | 90,52132701 | 34 | 90,28436019 | 34 |
| 12. | 91,23222749 | 34 | 90,52132701 | 38 |
| 13. | 85,78199052 | 33 | 86,7298582 | 33 |
| 14. | 88,62559242 | 34 | 88,15165877 | 37 |
| 15. | 86,492891 | 34 | 86,72985782 | 37 |
| 16. | 89,09952607 | 33 | 89,33649289 | 38 |
| 17. | 89,09952607 | 33 | 89,33649289 | 37 |
| 18. | 85,5450237 | 34 | 85,30805687 | 36 |
| 19. | 90,04739336 | 33 | 90,04739336 | 39 |
| 20. | 91,23222749 | 33 | 92,18009479 | 37 |
| Rata-rata | 88,27014218 | 33,55 | 88,36492891 | 35,5 |
| Deviasi standar | 2,012933724 | 0,998683344 | 1,865720243 | 2,328315406 |

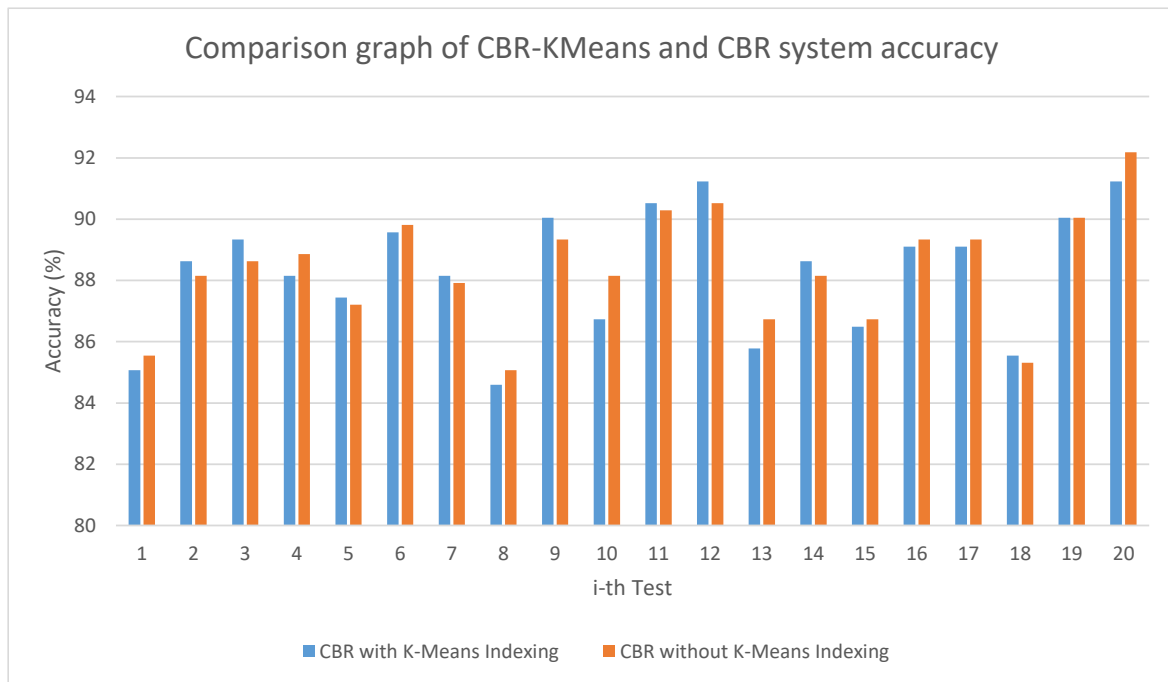


Fig 7. Comparison chart of the value of accuracy from CBR-K-Means and CBR system for 20 iterations

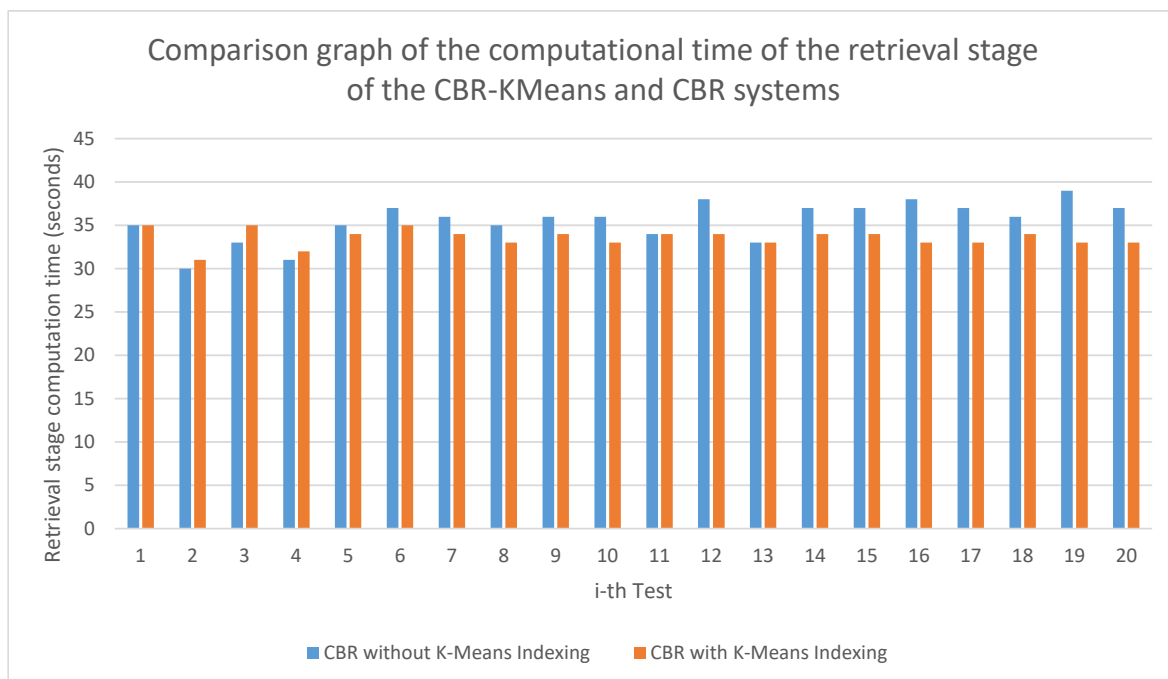


Fig 8. Comparison chart of the computational time for retrieval stage of the CBR-Kmeans and CBR system for 20 iterations

CONCLUSION

Conclusions of this study are:

1. This study resulted in a CBR system to estimate obesity rate by calculating the similarity value between new cases and old cases on a case base using *cosine similarity* and *euclidean distance*.
2. The test results of the test cases show that the CBR system without K-Means indexing method produces an average accuracy value of 88.365% and is better than the CBR system with K-Means

indexing method of 88.270% for 20 iterations at a threshold of 0.8.

3. The test results of the test cases show that the CBR system with K-Means indexing method produces an average computation time at retrieval stage of 33.55 seconds and is faster than the CBR system without K-Means indexing method of 35.5 seconds. This result is obtained with the number of clusters equal to two, the threshold is 0.8, and the test is carried out 20 iterations.

REFERENCES

- [1] Kementerian Kesehatan RI, "Panduan Pelaksanaan Gerakan Nusantara Tekan Angka Obesitas (GENTAS)," *Http://P2Ptm.Kemkes.Go.Id/Dokumen-Ptm/Panduan-Gentas*. pp. 6–16, 2017.2]
- [2] D. E. Safitri and N. S. Rahayu, "Determinan Status Gizi Obesitas pada Orang Dewasa di Perkotaan: Tinjauan Sistematis The Determinants of Urban Adults ' Obesity: Systematic review," *Arkesmas*, vol. 5, no. 1, pp. 1–15, 2020.
- [3] Kementerian Kesehatan RI, "FactSheet Obesitas Kit Informasi Obesitas.," *Jurnal Kesehatan*. pp. 1–8, 2018.
- [4] P2PTM Kemenkes RI, "Klasifikasi Obesitas setelah pengukuran IMT," Nov. 07, 2018. <http://p2ptm.kemkes.go.id/infographic-p2ptm/obesitas/klasifikasi-obesitas-setelah-pengukuran-imt> (accessed May 05, 2021).
- [5] F. M. Palechor and A. de la H. Manotas, "Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico," *Data Brief*, vol. 25, pp. 1–5, 2019, doi: 10.1016/j.dib.2019.104344.
- [6] A. Natassya, T. Tursina, and A. S. Sukamto, "Case Based Reasoning Diagnosis Risiko Penyakit Stroke Menggunakan Metode Hamming Distance," *J. Sist. Dan Teknol. Inf. JUSTIN*, vol. 7, no. 2, p. 122, 2019, doi: 10.26418/justin.v7i2.29891.
- [7] I. W. Supriana and K. D. Prebriana, "METODE PENALARAN BERBASIS KASUS (CASE BASE REASONING) DALAM PENENTUAN KELAYAKAN SEKOLAH PERAWATAN," *J. Resist. Rekayasa Sist. Komput.*, vol. 3, no. 1, pp. 57–65, 2020, doi: <https://doi.org/10.31598>.
- [8] N. W. Wiantari and I. W. Supriana, "Implementation of the CBR (Case-Based Reasoning) Method in Cases of Caesarean section," vol. 8, no. 2, pp. 181–189, 2019.
- [9] E. Wahyudi and S. Hartati, "Case-Based Reasoning untuk Diagnosis Penyakit Jantung," *IJCCS Indones. J. Comput. Cybern. Syst.*, vol. 11, no. 1, p. 1, 2017, doi: 10.22146/ijccs.15523.
- [10] H. Santoso and A. Syafrianto, "CASE BASE REASONING UNTUK MENDIAGNOSIS PENYAKIT HIPERTENSI MENGGUNAKAN METODE INDEXING DENSITY BASED SPATIAL CLUSTERING APPLICATION WITH NOISE (DBSCAN)," *Ethos J. Penelit. Dan Pengabd. Kpd. Masy.*, vol. 7, no. 1, pp. 88–100, 2019, doi: <https://doi.org/10.29313/ethos.v7i1.4206>.
- [11] U. S. Aesy and R. Wardoyo, "Prediction of Length of Study of Student Applicants Using Case Based Reasoning," *IJCCS Indones. J. Comput. Cybern. Syst.*, vol. 13,

- no. 1, p. 11, 2019, doi: 10.22146/ijccs.28076.
- [12] N. Rumui, A. Harjoko, and A. Musdholifah, "Case-Based Reasoning for Stroke Disease Diagnosis," *IJCCS Indones. J. Comput. Cybern. Syst.*, vol. 12, no. 1, p. 33, 2018, doi: 10.22146/ijccs.26331.
- [13] Salamun, "Penerapan Algoritma Nearest Neighbor dan CBR pada Expert System Penyimpangan Perilaku Seksual," *JOIN J. Online Inform.*, vol. 2, no. 2, pp. 63–70, 2017, doi: 10.15575/join.v2i2.97.
- [14] Suyanto, *Machine Learning Tingkat Dasar dan Lanjut*. Bandung: Informatika Bandung, 2018.
- [15] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*, Revisi. Bandung: Infomatika Bandung, 2018.
- [16] N. Chamidah, Wiharto, and U. Salamah, "Pengaruh Normalisasi Data pada Jaringan Syaraf Tiruan Backpropagasi Gradient Descent Adaptive Gain (BPGDAG) untuk Klasifikasi," *J. ITSMART*, vol. 1, no. 1, pp. 28–33, 2012, doi: 10.20961/its.v1i1.582.
- [17] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *CESS J. Comput. Eng. Syst. Sci.*, vol. 4, no. 1, pp. 78–82, 2019, doi: 10.24114/cess.v4i1.11458.
- [18] M. Salmin and S. Hartati, "Case Based Reasoning untuk Diagnosis Penyakit Infeksi Saluran Pernapasan Akut," *JIKO J. Inform. Dan Komput.*, vol. 1, no. 1, pp. 21–26, 2018, doi: 10.33387/jiko.v1i1.1167.
- [19] B. Purnama, *Pengantar Machine Learning Konsep dan Praktikum Dengan Contoh Latihan Berbasis R dan Python*. Bandung: Infomatika Bandung, 2019.
- [20] N. M. Nawi, M. Makhtar, Z. A. Afip, M. Z. Salikon, and K. A. M. Sukri, "Parkinson disease classification: A comparative analysis on classification techniques," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 3, pp. 1351–1358, 2020, doi: 10.11591/ijeecs.v18.i3.pp1351-1358.