JSCE (Journal of System and Computer Engineering)
Vol.6, No.1, Januari 2025, pp. 25-34
ISSN (online): 2723-1240
DOI : https://doi.org/10.61628/jsce.v6i1.1532

Clusterization Of Infant Data Based on Posyandu Examination Using K-Means

Muh.Ramli [1,a,*]; Sri Wahyuni [2,a]; Slamet Rayadi [3,a]
[1] Institut Teknologi & Bisnis Muhammadiyah Polewali Mandar, Jl. Poros Majene - Mamuju, Polewali Mandar 91352, Indonesia
[2,3] Institut Teknologi & Bisnis Muhammadiyah Polewali Mandar, Polewali, Polewali Mandar 91352, Indonesia
[a] mhdraml1y@gmail.com; [b] sriwahyuni@itbmpolman.ac.id; [3] slamet@itbmpolman.ac.id
** Corresponding author*

Abstract

Nutrition is a very important aspect for the human body, especially in toddlers and children. Balanced nutrition not only supports children's growth and development, but also improves academic achievement and contributes positively to their future development. However, in Tumpiling Village, the problem faced is the low basic understanding of parents and Posyandu cadres regarding balanced nutrition in early childhood. This causes toddlers to still be found with malnutrition or obesity, as well as a lack of data collected based on children's nutritional characteristics. Clustering, as one of the popular methods in processing medical, biometric, and various other fields of data, is known for its simplicity and effectiveness in grouping large-scale data based on similar characteristic. This study aims to groups the nutritional status of toddlers based on height and weight parameters using the K-Means Clusterings algorithm. This grouping produces several categories of nutritional status, namely obesity, overnutrition, good nutrition, undernutrition, and poor nutrition. By applying the Clustering method using K-Means, the nutritional status of toddlers can be classified more clearly, so that it can be a basis for Posyandu cadres in taking early preventive measures against malnutrition and obesity. In this study, the author used 28 toddler data. From the data, the author randomly determined the cluster center of 5 data, which then resulted in the following grouping: 7 toddlers experienced malnutrition, 3 toddlers were undernourished, 6 toddlers with good nutrition, 7 toddlers were overnourished, and 5 toddlers were obese. These results indicate the need for further attention and action from Posyandu and Puskesmas cadres to help parents in overcoming toddler nutrition problems.

Keywords— *K-Means, Clustering, Baby, Posyandu, Website.*

1. Introduction

Child development is one of the most important things that parents focus on, especially regarding the importance of achieving balanced nutrition. Currently, Indonesia is still facing infant nutrition problems, namely undernutrition and overnutrition (Setiawan et al., 2023). According to UN data, in 2020, more than 149 million children under the age of five in the world suffered from stunting, including 6.3 million from Indonesia, due to chronic malnutrition that can disrupt child development. On the other hand, widespread overfeeding is also showing an increase(Abdullah et al., 2023).

The principle of balanced nutrition includes the habit of consuming a variety of foods that are tailored to the needs of each family member. It is also important to adopt a regular healthy lifestyle, such as regular physical activity and monitoring your weight (Huang et al., 2023). A toddler's diet greatly affects his or her nutritional status. Optimal nutrition plays an important

role in supporting normal growth, intelligence and physical development in children. With good nutrition, the body becomes more resistant to diseases, infections, and is protected from the risk of chronic diseases (Widyaningsih, T. A., & Pratama, 2023).

Malnutrition is still one of the problems experienced by young children in Indonesia. The role of Posyandu and Puskesmas is very important in reducing the rate of malnutrition in children under five. Providing balanced nutritional intake at the right time is very important so that children can grow healthy and get used to living a healthy lifestyle in the future (Zhang, 2023). Nutritious food does not have to be expensive, so it is important that parents play a role in the selection of appropriate foods to meet the nutritional needs of toddlers. Grouping nutritional status in toddlers can be done with the K-Means method, which divides them into 5 clusters: undernourished, malnourished, overnourished, well-nourished, and obese (Fahrozi et al., 2023).

The nutritional status of infants can be determined based on weight for age (BB/u), weight for height (BB/TB), and height for age (TB/u), taking into account data on height, weight and gender. One of the methods used in assessing children's nutritional status is anthropometry. Anthropometric measurements, which are usually based on TB/u and BW/u, are often used to assess the nutritional status of children under five because this method is simple, safe, and does not require experts in the measurement stage (Unique, 2016). The results of anthropometric measurements are accurate and can detect a child's past nutritional history. In the research, this nutritional status grouping can be done using the Data Mining Clustering method with the K-Means technique, based on the child's height & weight (Suparto et al., 2022).

In anticipation of an increase in infant malnutrition cases, efforts are needed to improve information and understanding of proper parenting and diet. Active participation from parents as well as officers at the Posyandu is essential in monitoring the nutritional status of under-fives and ensuring children get a balanced intake and appropriate care to support their growth and development (Watrianthos & Suryadi, 2023).

Measuring a baby's height is still rarely done by parents, and if it is done, it often does not follow the right standards. In fact, by monitoring the child's height and weight regularly, the child's nutritional condition can be more easily monitored. If parents do not have the proper tools for such measurements, they should regularly visit Posyandu, which is conducted monthly by the local village, so that the child's nutritional status can be monitored properly (Zhou & Sun, 2024).

Posyandu is a basic service activity carried out by the local community, with support from health workers in the Puskesmas area. Generally, Posyandu is held regularly every month. In addition to providing child and maternal health services, Posyandu also monitors the growth and development of young children. Here, babies will be weighed and their height measured regularly, so that the development of their nutritional values can be monitored more accurately and precisely(Tumbelaka et al., 2018).

Therefore, data management was conducted at several Posyandu in Tumpiling Village, where the data was then grouped according to the nutritional characteristics of each child through cluster result analysis. This clustering was done based on 2 main parameters, namely toddler height (TB) and toddler weight (BW). The clustering method used is the K-Means algorithm, one of the effective and popular data mining techniques, especially in grouping large data into groups that have similar characteristics.

2. Method

In this study, the focus is on Posyandu cadres in Tumpiling Village, in monitoring growth and development in early childhood on a term basis. The data used include Posyandu information, Health Card (KMS), and measurements of height (TB) and weight (BB) in early childhood 0 to 36 months. The data processing model applied in this research is data mining with the K-Means method to classify toddler data. The data to be grouped is the data of toddlers Posyandu in Tumpiling Village taken in 2024. For data processing, various tools are used such

as Microsoft Excel, Rapid Miner, and website platforms that support data analysis. With this approach, it is hoped that a clearer picture of the nutritional status of children under five in the area can be obtained and support appropriate decision-making in efforts to improve child nutrition.

2.1 Clustering

Cluster is a group of data objects that are similar to each other in the same group, while objects in other groups have significant differences (Sen et al., 2023). Clustering, or cluster analysis, is the process of grouping a group of things both concrete and abstract into similar categories. The purpose of this process is to identify structure in data and group objects that share similar characteristics, making it easier to analyze and understand complex data. With this technique, we can discover previously unseen patterns or relationships in large datasets (Hastari et al., 2023)(Edastama et al., 2021)(Ren et al., 2022).

2.2 Data Mining

Data mining is a method used in finding interesting patterns and knowledge from large amounts of data (Papakyriakou, 2022). This process uses statistical, mathematical, machine learning, and artificial intelligence techniques to identify and extract functional information or knowledge from large databases. With data mining, we can uncover insights hidden in data, which can help in decision-making, trend analysis, and strategy development in various fields, including healthcare, business, and scientific research (Liu et al., 2023) (Edastama et al., 2021).

2.3 K-Means Clustering

K-Means is a data analysis method in Data Mining that operates with an unsupervised approach and is one of the techniques of data partitioning. In the K-Means method, data is grouped into several clusters, where each cluster has similar characteristics to each other, while differences with other clusters become significant. This method focuses on minimizing the variance between data in a cluster, while maximizing the difference between different clusters. In this way, K-Means helps identify structure in the data and facilitates further analysis based on the clusters formed. (Adam et al., 2023)(Munawar et al., 2021).

The K-Means method has several characteristics, including:

1) Simple and Easy to Use: K-Means is a relatively simple clustering method and can be easily applied, even by users who do not have an in-depth statistical or programming background.
2) Limitations in Segmentation: In certain types of datasets, K-Means may not be able to segment the data effectively. The clustering results sometimes do not reflect the cluster patterns that represent the natural features of the data, so it may produce irrelevant clusters.
3) Sensitive to Outliers: K-Means can run into problems when clustering data that has outliers. The presence of outliers can affect the cluster center (centroid), which in turn can lead to inaccurate clustering and reduce the quality of the clustering results. (Cordeiro et al., 2023).

In applying the K-Means algorithm, there are several things that must be considered, including:

a. Determining the number of clusters required
b. Only have attributes that are numeric:

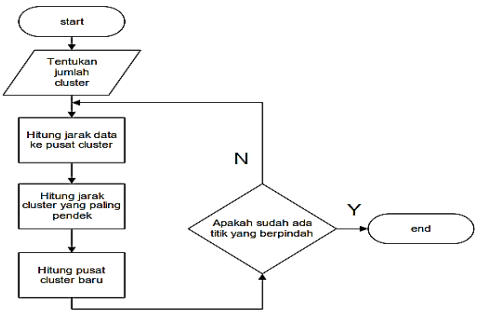


Figure 1. Flowchart of K-Means Clustering Algorithm

Figure 1 above shows the flow of the K-Means Clustering algorithm used in clustering the nutritional conditions of toddlers. The process in the K-Means algorithm can be explained as follows:

a. Determination of the number of clusters to be designated as cluster center k.
b. Use the Euclidean distance to calculate each data against the cluster center. Here is the Euclidean distance formula:

$$D(p,c)_n = \quad ^2 \qquad (1)$$

c. Categorize the data into clusters based on the shortest distance using the following equation:

$$\text{Min} \quad _{ik} \qquad (2)$$

d. Calculate the cluster center using the following equation:

$$C_{kj} = \qquad (3)$$

With: xij € kth cluster p = number of kth cluster members

e. Please repeat steps 2 to 4 until there is no more data moving to another cluster.

2.4 Population and Sample

The population used as the object of research consisted of 30 individuals selected using probability sampling techniques. in this population, the sample size was calculated using the Slovin formula as follows:

$$n = \frac{N}{1 + Ne^2} \qquad (4)$$

Dimana; n : sample size
N: population size
e: allowance for inaccuracy due to errors where:
n = 30 / (1 + (30) (0.0025)
n = 27.91 rounded to 28

Based on this formula, the number of samples obtained was 28 toddlers. The following is data on Body Weight (BB) and Height (TB) at the Tumpiling Village Posyandu:

Table 1. Under-five data.

Balita Ke-	Tinggi Badan (TB)	Berat Badan (BB)
Ke-1	120.6	13.8
Ke-2	116	9.7
Ke-3	119.6	13
Ke-4	103	14.3
Ke-5	129.1	17.2

Balita Ke-	Tinggi Badan (TB)	Berat Badan (BB)
Ke-6	108.3	10.2
Ke-7	84	10.2
Ke-8	81.5	10
Ke-9	83.5	10.7
Ke-10	103.5	15.4
Ke-11	87	10
Ke-12	126.6	14.9
Ke-13	83.5	96
Ke-14	77	9
Ke-15	88.2	12.3
Ke-16	72.2	8.2
Ke-17	92	11.3
Ke-18	124	13.1
Ke-19	70.7	7.5
Ke-20	65.2	6.3
Ke-21	58	5.1
Ke-22	130	18
Ke-23	110.5	13.1
Ke-24	122.3	13
Ke-25	105.7	18.2
Ke-26	123.4	15.5
Ke-27	120	13.4
Ke-28	114	10.9

3. Results And Discussion

The purpose of this research is to be able to carry out clustering of nutritional values of toddlers in Tumpiling Village using the K-Means method.

3.1 Normalization Calculation

The toddler data in Table 1 cannot be processed directly because there is a considerable difference in the numerical scale between the weight and height variables. The solution used in reducing the scale difference is to normalize the weight and height variables using the following formula :

Normalization Value =

Based on the variable values of weight and height will be normalized into the range of 0 to 1. This normalization process is required before calculating the centroid value in K-Means. The results of the normalization value calculation based on table 2 are as follows:

Table 2. Normalization value of toddler data

Balita Ke-	Tinggi Badan (TB)	Berat Badan (BB)
Ke-1	-57.51846154	-5.009375
Ke-2	-57.55384615	-5.052083333
Ke-3	-57.52615385	-5.017708333
Ke-4	-57.65384615	-5.004166667
Ke-5	-57.45307692	-4.973958333
Ke-6	-57.61307692	-5.046875
Ke-7	-57.8	-5.046875
Ke-8	-57.81923077	-5.048958333
Ke-9	-57.80384615	-5.041666667
Ke-10	-57.65	-4.992708333
Ke-11	-57.77692308	-5.048958333
Ke-12	-57.47230769	-4.997916667
Ke-13	-57.80384615	-4.153125
Ke-14	-57.85384615	-5.059375
Ke-15	-57.76769231	-5.025
Ke-16	-57.89076923	-5.067708333
Ke-17	-57.73846154	-5.035416667
Ke-18	-57.49230769	-5.016666667
Ke-19	-57.90230769	-5.075
Ke-20	-57.94461538	-5.0875
Ke-21	-58	-5.1
Ke-22	-57.44615385	-4.965625
Ke-23	-57.59615385	-5.016666667
Ke-24	-57.50538462	-5.017708333
Ke-25	-57.63307692	-4.963541667
Ke-26	-57.49692308	-4.991666667
Ke-27	-57.52307692	-5.013541667
Ke-28	-57.56923077	-5.039583333

3.2. Initial Cluster Center Value

After the weight and height data is normalized as in table 2, the next step is to determine the number of clusters or groups. The 28 toddler data in the previous table will be grouped into 5 clusters, namely: Undernutrition, Malnutrition, Good Nutrition, Obesity and Overnutrition. After the total clusters are determined, the next step is to determine the initial cluster center value for each cluster on each variable.

In determining the initial centroid, there are many methods that can be used. In this research, the method applied in data retrieval is by randomly selecting from the data source. The following is the determination of the Initial Cluster Center value that has been randomly selected:

Table 3. Initial Cluster Center Value

Cluster	Status Gizi	Tinggi Badan (TB)	Berat Badan (BB)

0	Obesitas	58.0	5.1
1	Gizi Lebih	129.1	17.2
2	Gizi Baik	92.0	11.3
3	Gizi Kurang	119.6	13.0
4	Gizi Buruk	105.7	18.2

3.3. Implementation of K-Means Clustering Rapid Miner Tools

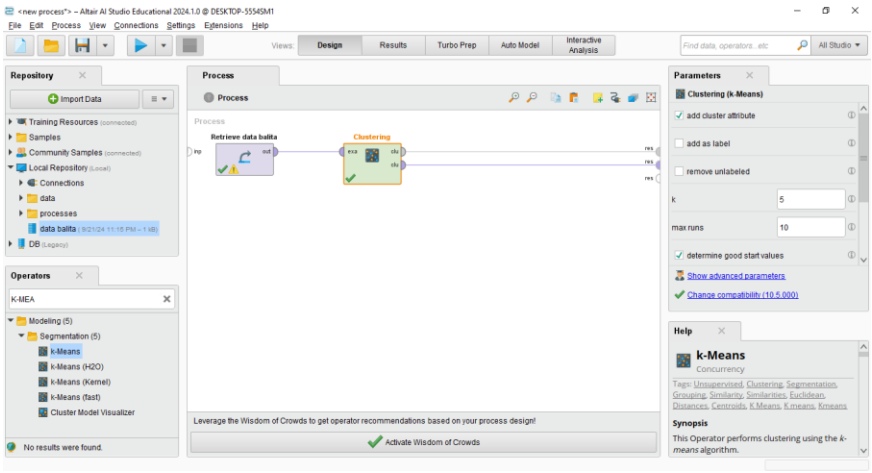
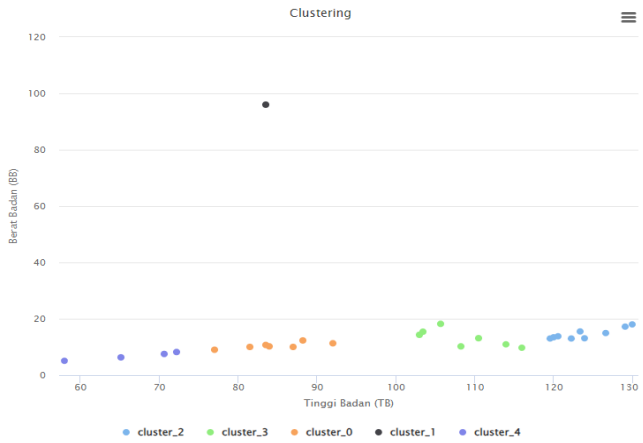


Figure 2. Clustering Process Design

clustering process design allows users to group data effectively.



Gambar. 3 Hasil chart *Clustering*

Chart results in the clustering process in RapidMiner to understand the structure and performance of the clusters that have been generated.

3.4. Infant Data Clustering Website Interface Design

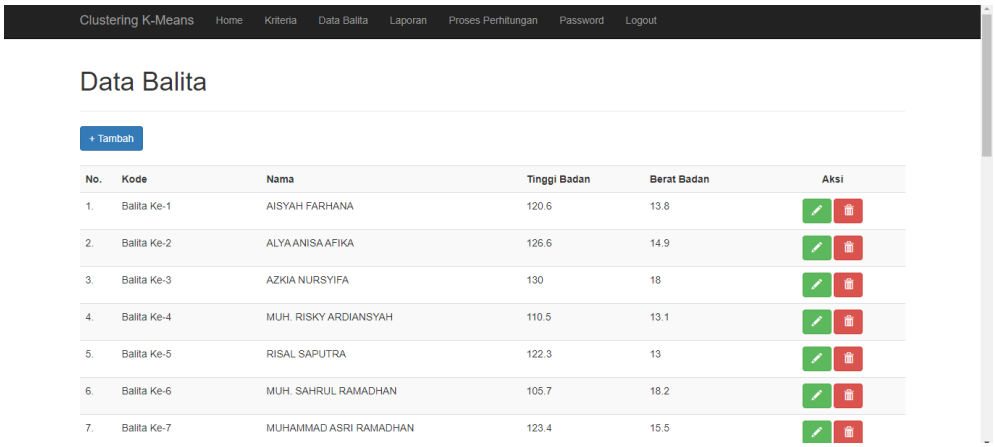


Figure. 4. Clustering data input for toddlers

In the figure above, regarding the results of the calculation of toddler nutrition using the K-Means Clustering method, it is found that the K-Means Clustering algorithm in grouping the nutritional conditions of toddlers at the Tumpiling Village Posyandu has results that are consistent with calculations made using Rapid Miner. Of the 28 existing toddler data, researchers determined the cluster center randomly as much as 5 data, which then resulted in the following groupings: 7 toddlers are malnourished, 3 toddlers are malnourished, 6 toddlers are well-nourished, 7 toddlers are overnourished, and 5 toddlers are obese.

4. Conclusions

In the discussion in the research on the implementation of K-Means Clustering in determining the nutritional conditions of toddlers at the Tumpiling Village Posyandu, researchers concluded that manual calculations using the K-Means Clustering algorithm in classifying the nutritional conditions of toddlers at the Tumpiling Village Posyandu have the same results as calculations carried out using Rapid Miner tools. Of the 28 existing toddler data, researchers determined the cluster center randomly as much as 5 data, which then resulted in the following groupings: 7 toddlers are malnourished, 3 toddlers are malnourished, 6 toddlers are well-nourished, 7 toddlers are overnourished, and 5 toddlers are obese. Therefore, support is needed from the puskesmas and related posyandu for baby parents, so that the number of malnourished toddlers can decrease in each coming year.

References

Abdullah, K. K., Maulana, I., Suharso, A., Primajaya, A., Komputer, F. I., Karawang, U. S., Barat, K. J., & Karawang, R. K. (2023). *PENERAPAN ALGORITME K-MEANS DALAM KLASTERISASI DATA*. *7*(2), 1274–1279.

Adam, I., Yenita, N., Sintuya, H., & Setthapun, W. (2023). Anomaly detection using K-Means and long-short term memory for predictive maintenance of large-scale solar (LSS) photovoltaic plant ☆. *Energy Reports*, *9*(S12), 154–158. https://doi.org/10.1016/j.egyr.2023.09.159

Cordeiro, R., Amorim, D., & Makarenkov, V. (2023). Neurocomputing On k -means iterations and Gaussian clusters. *Neurocomputing*, *553*(May), 126547. https://doi.org/10.1016/j.neucom.2023.126547

Edastama, P., Dudhat, A., & Maulani, G. (2021). Use of Data Warehouse and Data Mining for Academic Data : A Case Study at a National University. *International Journal of Cyber and IT Service Management*, *1*(2), 206–215. https://doi.org/10.34306/ijcitsm.v1i2.55

Fahrozi, A. Al, Insani, F., Budianita, E., & Afrianty, I. (2023). Implementasi Algoritma K-Means dalam Menentukan Clustering pada Penilaian Kepuasan Pelanggan di Badan Pelatihan Kesehatan Pekanbaru. *Indonesian Journal of Innovation Multidisipliner Research*, *1*(4), 474–492. https://doi.org/10.31004/ijim.v1i4.53

Hastari, D., Nurunnisa, F., Winanda, S., & Aprillia, D. D. (2023). *Application of K-Means and K-Medoids Algorithms for Grouping Country Data Based on Socio-Economic and Health Factors Penerapan Algoritma K-Means dan K-Medoids untuk Mengelompokkan Data Negara Berdasarkan Faktor Sosial-Ekonomi dan Kesehatan*. 274–281.

Huang, Q., Chen, S., & Li, Y. (2023). *Case Studies in Construction Materials Selection of seismic noise recording by K-means*. *19*(July), 1–16. https://doi.org/10.1016/j.cscm.2023.e02363

Liu, G., Ji, F., Sun, W., & Sun, L. (2023). ScienceDirect Optimization design of short-circuit test platform for the distribution network of integrated power system based on improved K-means clustering. *Energy Reports*, *9*, 716–726. https://doi.org/10.1016/j.egyr.2023.04.319

Munawar, A., Gendalasari, G. G., Kurniawan, I. M. G. A., Purnomo, D., Ependi, N. H., Rulinawaty, Indrawan, M. I., & Sadri, M. (2021). Cluster Application with K-Means Algorithm on the Population of Trade and Accommodation Facilities in Indonesia. *Journal of Physics: Conference Series*, *1933*(1). https://doi.org/10.1088/1742-6596/1933/1/012027

Papakyriakou, D. (2022). *Data Mining Methods : A Review Data Mining Methods : A Review*. *January*. https://doi.org/10.5120/ijca2022921884

Ren, Y., Pu, J., Yang, Z., Xu, J., Li, G., & Pu, X. (2022). *Deep Clustering : A Comprehensive Survey*. 1–20.

Sen, S., Adhikari, A., Ansary, K., & Roy, S. (2023). *Clustering Technique for Analyzing Leadership Style of the Head of the Institutions*. *May*. https://doi.org/10.48175/IJARSCT-9806

Setiawan, K. E., Kurniawan, A., Chowanda, A., & Suhartono, D. (2023). ScienceDirect ScienceDirect Clustering models for hospitals in Jakarta using fuzzy c-means and Clustering models for hospitals in Jakarta using fuzzy c-means and. *Procedia Computer Science*, *216*(2022), 356–363. https://doi.org/10.1016/j.procs.2022.12.146

Suparto, T. A., Azizah, N. N., & Andriyani, S. (2022). *The Problems Affecting the Implementation of Posyandu Program : A Literature Review*. *6*(1), 22–28.

Tumbelaka, P., Limato, R., Nasir, S., & Syafruddin, D. (2018). *Analysis of Indonesia ' s community health volunteers (kader) as maternal health promoters in the community integrated health service (Posyandu) following health promotion training*. *February*. https://doi.org/10.18203/2394-6040.ijcmph20180462

Unique, A. (2016). *済無No Title No Title No Title*. *3*(0), 1–23.

Watrianthos, R., & Suryadi, S. (2023). *Distribusi Spasial Unmet Need Pelayanan Kesehatan dengan Algoritma K- Means untuk Pemetaan Provinsi di Indonesia*. *4*(2), 361–368.

Widyaningsih, T. A., & Pratama, D. (2023). *PENERAPAN METODE K-MEANS CLUSTERING DALAM MENGELOMPOKKAN JUMLAH PESERTA BPJS KESEHATAN JKN / KIS*. *18*, 17–27.

Zhang, L. (2023). ScienceDirect Procedia ScienceDirect Research on K-means Clustering Algorithm Based on on K-means Clustering Algorithm Based on MapReduce Distributed Programming Framework MapReduce Distributed Programming Framework. *Procedia Computer Science*, *228*, 262–270. https://doi.org/10.1016/j.procs.2023.11.030

Zhou, Q., & Sun, B. (2024). Adaptive K-means clustering based under-sampling methods to solve the class imbalance problem. *Data and Information Management*, *June 2023*, 100064. https://doi.org/10.1016/j.dim.2023.100064