K-Means And K-Medoids Algorithms For Food Clusterization Optimized By Nutritional Value

Wildani Eko Nugroho[1*], Safar Dwi Kurniawan [2], Prayoga Alga Vredison[3], Hasbi Firmansyah[4]

[1,2,3] D3, Computer Enginering Program Study, Politeknik Harapan Bersama, Tegal, Indonesia
[3,4] Informatika, Universitas Pancasakti, Tegal, Indonesia
[*]Corresponding Author:
Email: wild4n1@gmail.com

Abstract.

Water, minerals, vitamins, carbs, proteins, and fats are the six categories of important elements that must be present in the food that people eat on a daily basis. Humans require nutrition in order to be able to do daily duties and be healthy. This study optimizes the K-Means and K-Medoids algorithms to employ the clustering method. Putting foods with comparable nutritional benefits together is the goal of this study. which foods are categorized into three groups: foods with high, medium, and low nutritional content. Finding the ideal number of clusters in data clustering is one of the difficulties. For this, the Elbow technique and the Davies-Bouldin Index (DBI) are frequently employed. The DBI evaluates the quality of clusters by taking into account the distance between clusters and the proximity between points inside the cluster, whereas the Elbow technique plots the intra-cluster variance versus the number of clusters to assist in determining the number of clusters. Better clusters are indicated by lower DBI values. The K-Means algorithm with the elbow approach produced results with a DBI of 0.643 and a cluster value (k) = 3 of 0.033. Regarding the K-Medoids algorithm using the Elbow approach, it has a DBI of 0.631 and a cluster value (k)=3 of 0.046. The optimal algorithm will be selected for clustering out of these two. According to the comparative results of the two methods, the K-Means approach yielded the best results for determining the number of clusters obtained, with a number of 0.033, while the K-Medoids method produced the best cluster, with a DBI value of 0.631. The study's clustering results can be utilized to choose and consume foods that will meet nutritional needs and help delay the onset of food-related disorders. For example, if you want to put on weight, you can choose foods in cluster 0. Cluster 2 foods can be picked if you wish to diet or lose weight, while Cluster 1 meals can serve as a benchmark if taken in excess, as this can lead to obesity. This conclusion may be seen from the clustering results, which demonstrate that foods with relatively high protein and calorie contents and low fat and carbohydrate content make up the first cluster. Foods with low carbohydrate content and high calorie, protein, and fat content make up the second cluster. Foods with minimal calorie, protein, fat, and carbohydrate content make up the third cluster.

Keywords: *K-Means; K-Medoids; Nutrition and Clustering.*

I. INTRODUCTION

One of the most important resources for human survival is food. Indonesia has a wide variety of cuisines. For instance, there are gudeg in Yogyakarta, serabi cake in West Java, spring rolls in Central Java, empek-empek in Palembang, and many other kinds of food. These foods differ in their flavors, textures, and other characteristics.The food that people eat on a daily basis must have six types of vital nutrients: water, vitamins, minerals, carbs, proteins, fats, and fats. For humans to carry out everyday duties and be healthy, they require nutrition [1]. Humans can obtain nutrition from plants or animal products. Animal meat contains protein, fat, carbs, and a few other nutrients like vitamins and minerals. In addition to animal flesh, eggs can provide the body with the nutrients it requires because they are composed of 75% water, 12% protein, and 12% fat [2]. Along with exercise and a healthy lifestyle, nutritional nutrition has a big impact on human health. Human needs must be met by food-based nourishment. If food has enough nutrients, it will be most useful and assist in increasing the body's metabolism [3]. But eating too much can also result in major health problems like high blood pressure, cancer, heart disease, stroke, and obesity. These diseases are the world's biggest cause of death, with dietary factors responsible for 70% of fatalities worldwide [4]. The Balance Principle One effective strategy to get past this is through nutrition. By altering lifestyles and taking into account the nutritional value of food in relation to the body's requirements, this theory is used to manage body weight and promote health [5].

It is evident from the discussion above that there hasn't been much work done to cluster or categorize foods according to their nutritional worth. K-Means and K-Medoids are two clustering algorithms that can be used to compute probabilities. This method works well for organizing data into many clusters [6]. One clustering technique that falls under the unsupervised learning category is the K-means algorithm. It divides data into multiple groups using the partition principle, and each group has a centroid that acts as its representation [6]. Some changes to the traditional K-Means algorithm may also be included in this approach, including data reduction, data weighting, and the application of the triangle inequality principle to expedite the nearest neighbor search [7].Additionally, the K-Means approach is capable of many evaluations. Internal procedures are employed for the evaluations. Evaluation is done internally by examining the density of the clusters and the distance between them [8]. The K-Medoids algorithm is the next technique used in this study. The K-Medoids algorithm is used in this study due to its excellent accuracy, ability to handle data that is sensitive to outliers, and efficiency in processing a large number of objects. The Euclidean distance Davies-Bouldin Index is used to evaluate the K-Medoids calculation results. The DBI value of 0.93543 indicates that the k-medoids algorithm produces good clustering because the calculation's end result is smaller than 0.93543 [9].

Similar studies were conducted using Elbow, DBI, and Silhouette optimization to compare K-Means and K-Medoids algorithms based on the locations of disaster-prone areas in Indonesia. Finding trends and improving catastrophe management techniques are the goals of this study. Geographical and historical data from several Indonesian disaster events, including Aceh Besar, Asahan, Badung, Bangkalan, Bekasi, and others, are included in the data. The ideal number of clusters is determined throughout the clustering process using optimization approaches like the Elbow Method, Davies-Bouldin Index (DBI), and Silhouette Score. With a Davies-Bouldin Index (DBI) result of 0.3737248981 and cluster 7, a Silhouette Score result of 0.868728638 and cluster 2, and an Elbow result of 98106477130.371 and cluster 2, the results demonstrate that the K-Means algorithm is generally more stable when handling outliers than K-Means. Analysis reveals an ideal site with the ideal amount of clusters [10].In light of this, scientists are interested in studying how to categorize food according to its nutritional worth, including its calorie, protein, fat, and carbohydrate content. This has already been accomplished, but only with the K-Means algorithm, which groups foods according to their nutritional value [3]. In addition to facilitating food selection and consumption for nutritional fulfillment, this clustering may help avert food-related illnesses. This study focuses on optimizing the K-Means and K-Medoids algorithms for food grouping according to nutritional value. By taking into account the nutritional value of food in relation to what the body requires, this research also seeks to alter lifestyles.

II. METHODS

2.1. Clustering

A method to divide a collection of data into groups according to a desired match is called clustering. In data mining clustering is the process of grouping data or objects into clusters to create data that is almost identical to the original and can be distinguished from items in other clusters [11].Clustering is used to separate data into regions with similar features such that each group has unique attributes. Without the use of preset labels or rules, clustering is a directed, or "unsupervised," data mining technique that arranges and searches data according to similarities between the data.

2.2. K-Means

K-Means is a clustering technique that falls within the unsupervised learning category. It divides into groups using the partitioning concept. The centroid is the representative of each group [6]. Organizing K partitions, also known as the centroid or mean of a particular data set, is the fundamental principle behind the K-Means algorithm.When n data is divided into k clusters using the non-hierarchical clustering technique K-Means, the Within-Cluster Sum of Squares is small, indicating strong cluster similarity, and the Between-Cluster Sum of Squares is large, indicating poor cluster similarity [3]. The steps in the K-Means algorithm method are as follows:

1) Calculate how many clusters need to form.
2) Using random selection, find each cluster's initial cluster center, or centroid.

3) Using the Euclidean distance in equation (1), get the distance between each observational data point and the chosen centroid.

$$dist(x,y)\sqrt{\sum_{i=1}^{\pi}(x_i - y_i)^2} \quad (1)$$

4) Grouping all observational data according to their minimum distance or proximity to the centroid.
5) The new centroid result value derived from the relevant cluster mean should be updated.
6) Continue steps 3 through 5 until each cluster's members remain the same.

2.3. K-Medoids

A partition clustering technique for forming clusters from a set of n objects is the K-Medoids algorithm. With this method, a specific item from the data set serves as a cluster center or representative. This algorithm's primary idea is to reduce the disparity between the objects and the selected cluster centers and make sure that the reference points in each cluster match the data that is provided [12]. The steps in the K-Medoids algorithm method are as follows:

1) Calculate how many clusters (k) need to form.
2) Choose at random each cluster's initial cluster centers, or medoids.
3) Using equation (2)'s Euclidean distance range, allocate each observation item to the closest cluster.

$$d(x,y)\sqrt{\sum_{i-1}^{\pi}(x_i - y_i)^2; i = 1,2,3, \dots n} \quad (2)$$

4) Choose items at random from each cluster to serve as the new medoids.
5) Determine the separation between the new medoids and every object in each cluster.
6) Determine the distance value of the new total distance less the old total distance to determine the total deviation (S). To create a new set of k objects as medoids, swap the cluster data objects if S is less than 0.
7) To create a cluster and each cluster member, repeat Steps 4 through 6 until the medoids remain unchanged.

2.4. Elbow Methods

One popular technique for figuring out how many clusters should be formed during the clustering process is the elbow approach. To maximize or ascertain the ideal number of clusters for the clustering process to be executed, the elbow approach is utilized. Cluster cohesiveness and separation are calculated using the elbow approach [13]. Cluster separation gauges how distinct or disjointed one cluster is from another, and cluster cohesiveness gauges how tightly related the data in the cluster is. The sum of squares error (SSE) can be used to quantify cluster cohesiveness and cluster separation [14]. The following formula can be used to calculate SSE :

$$SSE = \sum_{k=1}^{k} \sum_{xi} |x_i - C_k|^2 \quad (3)$$

Description :

K = Cluster in C

X_i = Distance of the data on the i-th object

Ck = The i-th cluster's center

2.5. Davies Bouldin Index (DBI) Methods

One technique to guarantee the optimal number of clusters following the clustering process is Davies Bouldin Index (DBI) [15]. The DBI technique aims to minimize the distance between objects within a cluster and maximize the difference across clusters. Additionally, the DBI measure is used to evaluate the effectiveness of various clustering techniques; values near zero imply better clusters. The DBI algorithm divides the distance between the cluster centers by the average of the compactness values for each place [16]. An outline of the procedures to determine the DBI value is provided below [17].

1) Finding the Sum of Squares The attachment between members of a cluster, or how similar they are to one another, is measured by the SSW; the smaller the cluster, the more similar the members are to one another. To find the matrix, cohesion, and homogeneity, SSW is computed. According to the equation formula, cohesion is the relationship between group members in a single cluster.

$$SSW = \frac{1}{m_i}\sum_{j=1}^{m_i} d(x_j \,.\, c_i) \quad (4)$$

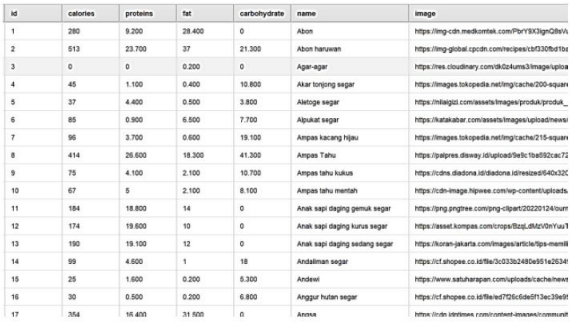
Description :

m_i = quantity of information in clusterke-i
x = information within the cluster
d(x,c) = data's distance from the centroid
x_i = information within the cluster
ci = center of cluster i centroid

2) To ascertain separation or heterogeneity, the Sum of Square Between Clusters (SSB), which is a sufficiently high distance between groups to keep them apart from one another, is calculated. The distinction between two groups is separation.
3) The purpose of calculating the ratio is to determine how well one cluster compares to other clusters. There must be more separations than cohesiveness.
4) DBI calculation

III. RESULT AND DISCUSSION

Pre-processing, also known as preparation, is the first step in the research process. This involves a number of steps, such as data preprocessing, which includes attribute selection, removing duplicate data, normalizing data, and lowering the dimensions of the dataset used. After undergoing data preprocessing, the final dataset consists of food nutritional values. On the final dataset, the K-means and K-medoids algorithms are used to get the optimal clusterization performance value. The K-Means and K-Medoids algorithms are used to evaluate the results during the testing phase. Figu 1. This is a screenshot of the Kaggle.com website's food nutrition dataset.



id	calories	proteins	fat	carbohydrate	name	image
1	280	9.200	28.400	0	Abon	https://img-cdn.medkomtek.com/PbrY9X3lgnQ8eVA
2	513	23.700	37	21.300	Abon haruwan	https://img-global.cpcdn.com/recipes/cbf330fbd1ba
3	0	0	0.200	0	Agar-agar	https://res.cloudinary.com/dk0z4ums3/image/upload
4	45	1.100	0.400	10.800	Akar tonjong segar	https://images.tokopedia.net/img/cache/200-square
5	37	4.400	0.500	3.800	Aletoge segar	https://mlilgtzl.com/assets/images/produk/produk_
6	85	0.900	6.500	7.700	Alpukat segar	https://katakabar.com/assets/images/uploadnewsi
7	96	3.700	0.600	19.100	Ampas kacang hijau	https://images.tokopedia.net/img/cache/215-square
8	414	26.600	18.300	41.300	Ampas Tahu	https://palpres.disway.id/upload/9e9c1ba092cac71
9	75	4.100	2.100	10.700	Ampas tahu kukus	https://cdns.diadona.id/diadona.id/resized/640x320
10	67	5	2.100	8.100	Ampas tahu mentah	https://cdn-image.hipwee.com/wp-content/uploads
11	184	18.800	14	0	Anak sapi daging gemuk segar	https://png.pngtree.com/png-clipart/20220124/ourn
12	174	19.600	10	0	Anak sapi daging kurus segar	https://asset.kompas.com/crops/BzqLdMzV0nYuuT
13	190	19.100	12	0	Anak sapi daging sedang segar	https://koran-jakarta.com/images/article/tips-memili
14	99	4.600	1	18	Andaliman segar	https://cf.shopee.co.id/file/3c033b2480e951e2634
15	25	1.600	0.200	5.300	Andewi	https://www.satuharapan.com/uploads/cache/news
16	30	0.500	0.200	6.800	Anggur hutan segar	https://cf.shopee.co.id/file/ed7f25c6de5f13ec39e9f
17	354	16.400	31.500	0	Angsa	https://cdn.idntimes.com/content-images/community

Fig 1. Food Nutrition Dataset on the Kaggle Site

A number of preprocessing procedures must be completed prior to the K-Means and K-Medoids algorithms being implemented. This is done in order to maximize the outcomes of using the K-Means and K-Medoids algorithms. Among the preprocessing procedures used in this study are the following ones:

a. Atribut selection

The purpose of attribute selection is to eliminate unnecessary qualities and choose those that will be utilized in the K-Means and K-Medoids algorithms. Since the K-Means and K-Medoids algorithms can only accept characteristics with numeric data types, unneeded attributes are removed. Therefore, non-numerical attributes must be chosen first. This is the outcome of choosing the traits that will be utilized later in Table 1.

Table 1. Result Of Attribute Selection

Nama Atribut	Keterangan
Calories	Kalori per 100 gram / makanan (kal)
Fat	Lemak per 100 gram / makanan (gram)
Protein	Protein per 100 gram / makanan (gram)
Carbohidrat	Karbohidrat per 100 gram / makanan (gram)

b. Remove Duplicate

By comparing each piece of data one at a time, the Remove Duplicates function eliminates duplicates. One data set from the comparison results will remain after all data with the same value are eliminated one at a time. Since the nutritional value of each product is typically the same, this research removes redundant data to increase the efficiency of data processing.

c. Outlier Detection

Finding data that behaves differently from other typical data requires outlier detection, which will have an impact on how the K-Means and K-Medoids algorithms analyze the data. Distance-based outlier identification is the outlier identification method employed in this study. Two points are determined by selecting a distance-based outlier, which is subsequently verified. It will be regarded as an outlier if the neighbor points are near together and then separated. Once this study has been tested, the parameters' findings are derived utilizing the distances of ten neighbors, nine outliers, and the Euclidean distance.

d. Normalization of Data

One of the data normalizing techniques employed in this study is min-max normalization. The original data values are transformed into various values between 0 and 1 using this min-max normalization. Forming data into numbers with the same range is the goal of the normalization process, which speeds up data processing.

e. The reduction of dimensions

Principal Component Analysis (PCA) is a technique used to minimize the dimensions of data containing numerous variables. PCA shrinks the data's dimensionality without appreciably lowering its information content. Based on the data, PCA creates a new set of dimensions that are rated. The principal components of PCA are derived from the breakdown of the covariance matrix's eigenvalues and eigenvectors. Two components and a fixed number of dimensions are the parameters utilized for PCA implementation. Two traits remained after the previous four were eliminated.

3.1. K-Means Algorithm

A number of tasks or procedures need to be completed in experiments utilizing the K-Means method, including dimension reduction, outlier detection, attribute selection, data standardization, and the elimination of duplicate data. The Elbow method and the Davies-Bouldin Index method are the two techniques used to test the K-Means method. Fig 2. shows the results of the K-Means experiment:

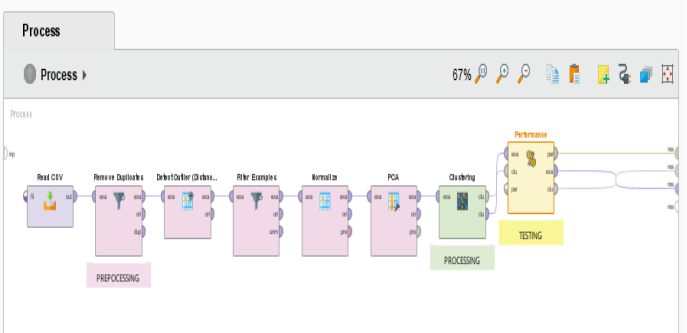


Fig 2. Steps for Testing the K-Means Algorithm

The K-Means algorithm's test display is shown in the image above. The Elbow method and the Davies-Bouldin Index method are the two approaches utilized in several of these exams. Testing is done to determine the K-Means algorithm's accuracy and success rate. The Elbow approach and the Davies-Bouldin Index method are two methods that can be used to determine the number of clusters.

a) Elbow Methods

A fundamental stage in all unsupervised algorithms is the Elbow Method, which establishes the ideal value for the K value in the K-Means algorithm by counting the number of clusters. Cluster values from 1 to n are run using the Elbow Method, and each value is computed using the Cluster Sum of Squares (WCSS). The sum of squares between each cluster's centroid value and average value is known as WCSS. Then use the chart line to determine which point produces the most elbow. Table 2 displays the outcomes of testing the Elbow method on the K-Means algorithm. K value and average. Figure 3: Graph of Cluster Number Determination for K-Means Algorithm and Within-Cluster Sum of Squares.

Table 2. K Values and Avg. Within-Cluster Sum Of Square K-Means

K	WCSS
2	0,063
3	0,033
4	0,025

5	0,019
6	0,016
7	0,013

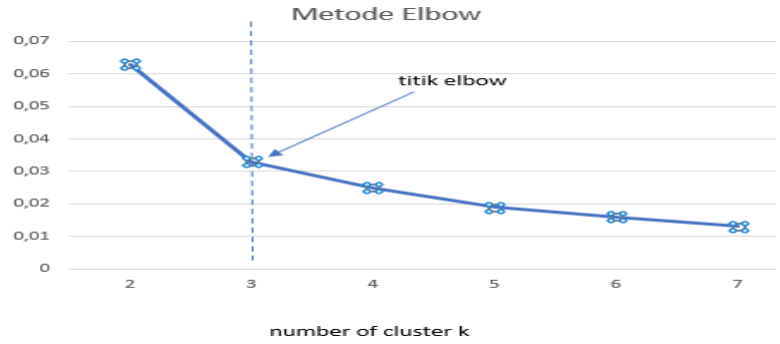


Fig 3. Calculating the K-Means Algorithm's Cluster Count

b) Davies Bouldin Index (DBI) Method

The Davies-Bouldin Index (DBI) is a cluster validation that integrates separation and cohesion statistics. Separation is the distance between cluster centers, whereas cohesion is the degree of resemblance between data and cluster centers. The quantity and proximity of the clustered data determine the quality of the clustering results. The cluster outcomes are better when the DBI value is smaller (non-negative => 0). The outcomes of applying the Davies-Bouldin Index method to the K-Means algorithm, as displayed in table 3, are as follows.

Table 3. K-Means Value for the Davies-Bouldin Index (DBI)

K	DBI
2	0,746
3	0,643
4	0,703
5	0,736
6	0,751
7	0,791

The test results of the Davies-Bouldin Index method are displayed in Table 3 above. The smallest number, 0.643, is the smallest value found in cluster k = 3. The optimal number of clusters is 3, according to the Elbow and Davies-Bouldin Index techniques that have been used. Table 4 below displays these findings:

Table 4. K-Means Value for WCSS

Cluster	Avg. Within Centroid Distance
0	0,063
1	0,038
2	0,024

Based on the data in Table 4, the average within-centroid distance value is near zero, indicating good data closeness within the cluster. A scatter plot can be used for simpler visualization.

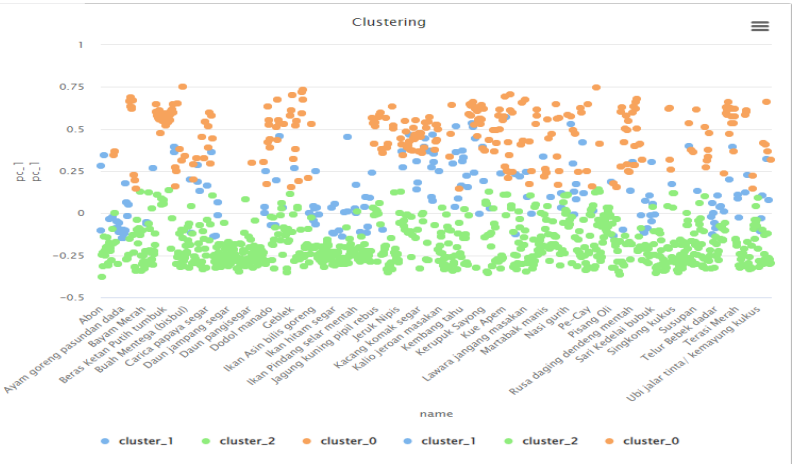


Fig 4. K-Means Algorithm Scatter/Bubble Outcome with Three Clusters

Following the K-Means method operation, the clusterization results are shown in table 5, which shows the quantity of data for each cluster, and table 6, which shows the food data that was produced as a result of clusterization.

Cluster	List Of Food	Descriotion
0	Asam masak di pohon, Asan kandis kering, Bagea kelapa, Bagea kenari, Bakwan, Bantal, Beras Cerdas, Beras ganyong, Beras giling varian pelita mentah, Beras hitam mentah, Beras jagung kuning kering mentah, Beras jagung putih kering mentah, Beras Ketan Hitam, Beras Ketan Hitam tumbuk, Beras Ketan Putih, Ikan Asin bilis goreng, Ikan Asin gabus goreng, Ikan Asin japuh goreng, Ikan Asin pari goreng, Ikan Asin pepetek goreng, Ikan Asin sirinding goreng, Ikan Asin teri goreng, Ikan bandeng presto masakan	According to this cluster, the dietary data had a high calorie content (250–400 calories), low protein and fat content (average <10 mg), and high carbohydrate content (over 40 mg) per 100 grams of food.
1	Abon, Abon Haruan, Ampas, Tahu, Angsa, Arwan Sisir, Ayam, Ayam goreng Kentucky sayap, Ayam ampela goreng, Ayam goreng church texas sayap, Ayam goreng church texas dada, Ayam goreng kalasan paha, Ayam goreng kentucky dada, Ayam goreng Kentucky paha, Ayam goreng mbok berek dada, Ayam goreng paha, Ayam goreng Pasundan dada, Ayam goreng Pasundan paha, Ayam goreng pioneer dada, Ayam goreng Sukabumi dada, Ayam goreng Sukabumi paha, Ayam hati segar, Ayam Taliwang masakan, Ayam usus goreng, Emping beras, Emping komak, Wijen, Worst (sosis daging)	According to this cluster, a 100-gram meal has more than 260 kcal of calories, more than 20 mg of protein on average, more than 10 mg of fat, and less than 5 mg of carbohydrates.
2	Agar-agar, Akar Tonjong segar, Aletoge segar, Alpukat segar, Ampas kacang hijau, Ampas tahu kukus, Ampas tahu mentah, Anak sapi daging gemuk segar, Anak sapi daging kurus segar, Anak sapi daging sedang segar, Andaliman segar, Andewi, Anggur hutan segar, Anyang sayur, Apel, Apel malang segar, Arbei, Ares sayur, Arrowroot segar, Asam kandis segar, Asam payak segar, Asinan Bogor sayuran, Ayam dideh/darah segar, Babi ginjal segar, Babi hati segar, Bacang, Baje, Bakso Bakung segar, Barongko, Batang Tading, Batar daan, Batatas gembili segar, Batatas kelapa ubi bakar, Batatas kelapa ubi kukus, Batatas kelapa ubi segar, Batatas tali ubi rebus, Bawang Bombay, Bawang Merah Bawang Putih, Bayam kukus, Bayam Merah, Bayam merah segar, Bayam rebus, Bayam segar Bayam tumis + oncom, Bayam tumis bersantan, Coto mangkasara kuda masakan, Coto mangkasara sapi masakan, Cue Selar Kuning, Cuka, Daun kenikir segar, Daun kesum segar, Daun kol sawi segar, Daun Koro, Daun kubis segar, Daun Kumak, Daun Labu Siam, Daun Labu Waluh, Daun lamtoro segar, Daun leilem segar, Daun Leunca, Daun Lobak Daun Lompong Tales, Daun Mangkokan, Sawo duren segar, Sawo kecik segar, Sayur asem, Sayur bunga papaya, Sayur garu, Sayur kohu-kohu, Sayur lebui, Sayur lilin-terubuk, Sayur ndusuk, Sayur sop	With calories below 100 kcal, protein and fat content averaging less than 10 mg, and adequate carbohydrate content averaging more than 10 mg per 100 grams of food, this cluster discovered that the meal data had inadequate nutritional quality.

Three groups, each with distinct outcomes, are identified using the K-Means Algorithm implementation results. Foods in Cluster 0 are heavy in calories and carbohydrates and low in fat and protein. Because calories and carbs provide the primary energy, foods in cluster 0 can be selected as

alternatives to staple foods. Furthermore, if ingested in adequate amounts, it can also be used to gain weight; nevertheless, if consumed in excess, foods in cluster 0 can lead to obesity.Foods in Cluster 1 are rich in calories, somewhat high in protein, and low in fat and carbohydrates. This allows you to consider meals in cluster 1 if you wish to eat them. The items in cluster 1 are high in calories, despite having a relatively high protein level. To acquire adequate protein, it is beneficial to eat it multiple times per week. On the other hand, obesity may result from overconsumption combined with insufficient exercise.Foods with poor nutritional content are included in Cluster 2. Due to their low calorie, carbohydrate, and fat content, foods in this cluster are suitable for dieting or weight loss. It can, however, cause the body to become weak and deficient in nutrients if it is not balanced with sufficient dietary intake.

3.2. K-Medoids Algorithm

A number of tasks or procedures need to be completed in experiments utilizing the K-Medoids algorithm, including dimension reduction, outlier detection, attribute selection, data standardization, and the elimination of duplicate data. The Elbow method and the Davies-Bouldin Index method are the two techniques used to test the K-Medoids algorithm. A perspective of the K-Medoids experiment is shown in Figure 2.

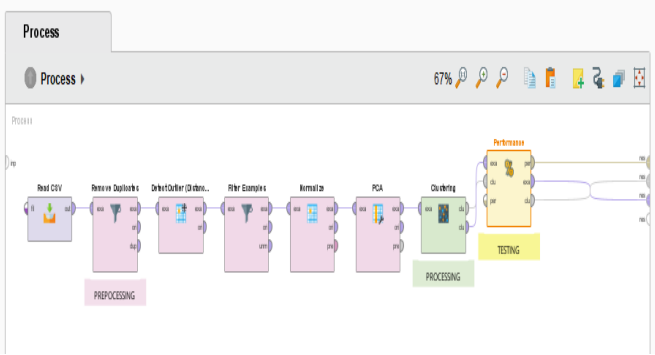


Fig 5. Algorithm Testing Scenario for K-Medoids

The K-Medoids method is tested in the image above. Both the DBI approach and the elbow method are employed in some of these exams. The K-Medoids algorithm is tested to determine its accuracy and success rate. The DBI approach and the elbow method are two methods that can be used to determine the number of clusters.

a) The Elbow Method

In order to determine the number of clusters in the K-Medoids algorithm, the foundation of this phase is to establish an ideal K value. Each value in the Cluster Sum of Squares (WCSS) will be computed using the Elbow technique, which runs cluster values from 1 to n. The WCSS represents the sum of squares between each cluster's centroid value and average value [18]. Next, using a chartline, choose the spot that creates the most elbow. The outcomes of applying the Elbow approach to the K-Medoids algorithm, as displayed in Table 5, are as follow. Within-cluster sum of squares, K value, and average Figure 3. Graph Calculating How Many Clusters the K-Medoids Algorithm Needs.

Table 5. Avg. Within-Cluster Sum Of Square K-Medoids

K	WCSS
2	0,091
3	0,046
4	0,036
5	0,027
6	0,023
7	0,019

The Avg. Within-cluster Sum of Square K-Medoids, which calculates the sum of these squared distances for each data point within each cluster, is displayed in the above table. The ideal number of clusters in K-Medoids is found using WCSS. The cluster is more compact and the data within the cluster is more uniform when the WCSS is smaller. To determine the ideal number of clusters, the aforementioned table is also utilized to test three to seven clusters. The Elbow Method cluster experiments are shown in Fig 6. 3-7.

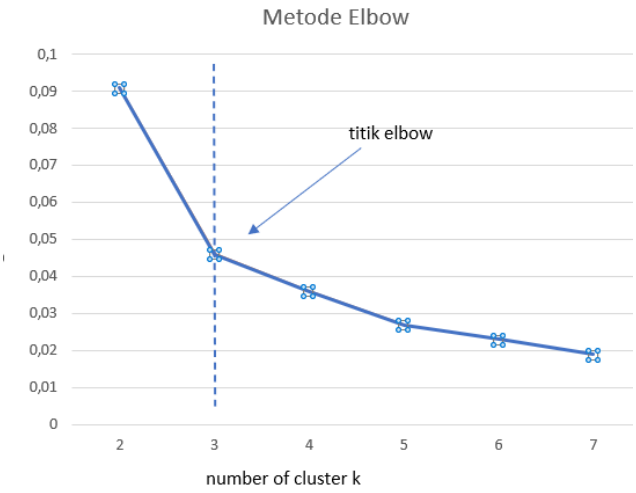


Fig 6. Calculating The K-Medoids Algorithm's Cluster Count

The outcome of an experiment to find the ideal number of clusters based on the SSE (Sum of Square Error) value on the K-Medoids algorithm using the Elbow technique is shown in Figure 6. In this instance, the SSE value has significantly dropped. The cluster's quality decreases with increasing SSE value and vice versa. The cluster quality improves with a reduced SSE value. As can be seen in the above image, the SSE value is at its peak when there are two clusters (k = 2), and it dramatically decreases but forms an elbow when there are three clusters (k = 3). The SSE value dropped once more when there were four clusters (k = 4), and so on until there were seven clusters (k = 7). The elbow is found at the number of clusters k = 4 by calculating the average value. This is because the graph shows that the number of clusters that form an elbow is clearly visible when the number of clusters k = 3, and the number of clusters k = 4 to k = 7 is seen to start decreasing as well. sum of squares within the cluster of 0.046.

b) Davies Bouldin Index (DBI) Method

A cluster validation that incorporates separation and cohesion data is the Davies-Bouldin Index (DBI). Cohesion is the degree of similarity between the data and the cluster centroids, whereas separation is the distance between the cluster centroids. The quantity and proximity of the clustered data determine the quality of the clustering results. The cluster outcomes are better when the DBI value is smaller (non-negative => 0). The outcomes of applying the Davies-Bouldin Index approach to the K-Medoids algorithm, as displayed in table 6, are as follows.

Table 6. Value of the Davies-Bouldin Index K-Medoids

K	DBI
2	0,810
3	0,631
4	0,996
5	1,009
6	0,996
7	1,144

According to the above table, cluster k = 3 has the minimum DBI value, which is 0.631. This demonstrates that the outcomes match expectations. The optimal number of clusters is three, according to the Elbow and Davies-Bouldin Index techniques that have been used. Table 7 below displays these findings:

Table 7. Value of WCSS K-Medoids

Cluster	**Avg. Within Centroid Distance**
0	0,091
1	0,046
2	0,036

As can be seen from the above table, the cluster's data are reasonably close to one another, as indicated by the Average Within Centroid Distance value being near zero. The scatter/bubble in Figure 7 below can be used for simpler display.

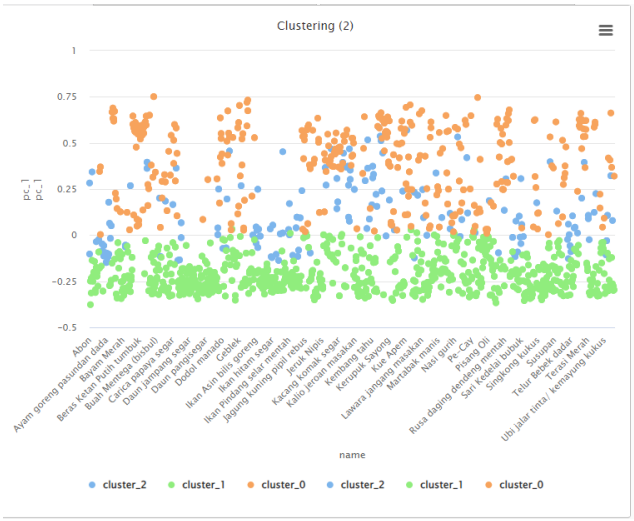


Fig 7. K-Medoids Algorithm Scatter/Bubble Outcome with Three Clusters

Following the K-Medoids algorithm operation, the clusterization results are shown in table 8, which shows the quantity of data for each cluster, and table 9, which shows the food data that was produced as a result of clusterization.

Table 8. The quantity of information for every K-Medoids cluster

Cluster	Jumlah Data
0	347
1	744
2	175

Table 9. Some Food Information from the Clustering Findings of K-Medoids

Cluster	List Of Food	Description
0	Asam payak segar, Bagea kelapa asin, Bagea kelapa manis, Bagea kenari asin, Bagea kenari manis, Bagea kw 1, Bagea kw 2, Bakpia kue, Bakwan, Bantal, Batatas tali ubi rebus, Beef burger, Belitung talas kukus, Beras ganyong, Beras Giling, Beras Giling masak (nasi), Beras ladang mentah, Beras Menir, Beras menir mentah, Beras Merah tumbuk, Bihun, Bihun goreng instan, Bihun Jagung mentah, Bihun mentah, Biji nangka/Biji salak, Bika ambon, Bika Ambon, Biskuit, Bonggol Pisang kering, Brem, Brondong, Buah kom segar, Bubur sagu, Bulung Jajan, Bulung Sangu, Bungkil Biji Karet, Bungkil Kelapa, Burung sarang segar, Cabai merah kering, Cake tape, Cantel mentah, Cengkeh kering, Centel, Ceriping getuk singkong, Coklat bubuk, Coklat Manis batang, Coklat Susu batang, Combro, Daun ndusuk segar, Daun salam bubuk, Deblo, Djibokum masakan, Dodol, Dodol bali, Dodol BanjarmasinEmping komak, Enting-enting gepuk hello kity, Es Krim (Coconut milk), Es Mambo, Gadeng/Gadung kering, Gaplek, Gatep segar, Gatot, Geblek, Gelatine, Gemblong, Gula putih, Gurandil, Havermout, Intip goreng, Jagung grontol, Jagung Kuning giling, Jagung Kuning pipil baru, Jagung uning pipil kering mentah, Jagung pipil var. harapan kering, Jagung pipil var. metro kering, Jagung Rebus, Jagung titi, Jali, Jam selai, Jampang huma mentah, Jamur encik, Jamur kuping kering, Japilus, Jawawut, Jenang, Jengkol segar, Kabuto, Kacang Andong, Kacang Arab, Kacang babi kering, Kacang Bogor goreng, Kacang tunis kering, Kacang uci kering, Kacang urei kering, Kambose, Katul Beras, Katul Jagung, Kecimpring singkong goreng, Kelepon, Kentang Hitam, Keremes, Keripik gadung, Kerupuk Aci, Kerupuk cumi goreng, Kerupuk Ikan berpati, Kerupuk kemplang (ikan) mentah, Kerupuk kemplang goreng, Ketumbar, Ketupat ketan, Kopi bagian yang larut, Kopi bubuk instant, Koro	This cluster found that the food data had a high calorie content (60–720 calories), low average levels of fat and protein (less than 50 mg), and high levels of carbohydrates (above 300 mg) per 100 grams of food.

Cluster	List Of Food	Description
	Benguk biji, Koro Roway biji, Koro Wedus biji, Koya, Koya mirasa, Kranji segar, Kue ali, Kue Apem, Kue bangket, Lopis, Lupis ketan, Madu, Maizena tepung, Makaroni, Markisa squash, Martabak manis, Martabak mesir, Martabak Telur, Masekat, Melase, Merica, Mie Bendo, Mie Goreng, Mie kering, Mie Sagu, Misoa, Misro, Nangka biji, Nasi, Nasi beras merah, Nasi Goreng, Nasi jagung, Nasi minyak, Nopia special, Oncom ampas kacang hijau, Onde-onde, Oyek, Oyek (dari singkong), Pastel, Pati Singkong (tapioka), Pempek belida, Pempek kulit, Pempek tenggiri, Pepaya lodeh, Permen, Petis Udang, Petis udang kering, Petis udang pasta, Pisang ketip segar, Pisang Mas, Pisang Oli, Pulut, Putu Mayang, Rarawuan, Rasbi (Beras Ubi), Rasi (Beras Singkong), Rebung laut mentah, Rempeyek kacang uci, Renggi goreng, Risoles, Roti boong, Roti Gambang, Roti Putih, Roti warna sawo matang, Sagu Ambon, Sale pisang siam. Santan (kelapa saja), Sapi abon, Sarimuka, Semur Jengkol, Sente talas segar, Serbuk Coklat, Serimping talas kebumen, Setrup sirup, Singkong Goreng, Singkong kukus, Singkong tape, Srikaya, Susu asam untuk bayi bubuk, Susu Kental Manis, Susu skim bubuk, Suwir-suwir, Talas pontianak segar, Tapai singkong, Tauco cap DAS cake, Tauji cap singa, Teh, Teh hijau daun kering, Teh melati daun kering, Tekokak kering, Tempe Sayur, Wajit camilan, Widaran, Wingko babat, Yangko	
1	Bekasam, Bekasang, Belibis daging segar, Belimbing, Belut segar, Belutlaut segar, Bengkuang, Bentul (Komba) talas segar, Bentul talas kukus, Betok wadi masakan, Bihun Goreng, Bir (4% alkohol), Bit, Biwah segar, Buah ruruhi segar, Buah tuppa segar, Bubur tinotuan (Manado), Bulgogi masakan, Buncis, Buncis asam, Buncis rebus, Bunga pepaya segar, Bunga turi segar, Buntil, Buntil daun talas, Buras, Cabai gembor merah segar, Cakalang asar (asap papua), Cammetutu, Cap cai sayur, Carica papaya segar, Cempedak, Dideh darah ayam, Dideh darah sapi, Dodonkol, Domba daging kurus segar, Domba ginjal segar, Duku, Durian, Duwet segar, Eceng Gondok, Embacang, Encung asam segar, Erbis, Es krim, Es Sirup, Fillet o-fish, Gado-gado, Gadung kukus, Gadung mentah, Gadung ubi kukus, Gadung ubi segar, Gambas (Oyong), Gambas lodeh, Gandaria, Ganyong kukus, Ganyong mentah, Ganyong segar, Gembili, Gembili ubi segar, Genjer segar, Gete kuah asam masakan, Ginjal Babi, Ginjal Domba, Ginjal Sapi, Gudeg, Gudeg sayur, Gulai asam keueung masakan, Hangop, Hati Babi, Hati Sapi, Hofa/Ubi hutan segar, Ikan Patin segar, Kentang, Keong, Kepala Susu (Krim), Kepiting, Kerang, Kerbau daging segar, Keribang ubi segar, Kerokot segar, Kesemek, Ketimun, Ketimun madura segar, Ketoprak, Ketupat Tahu, Kokosan, Komak polong segar, Koro wedus polong, Kotiu hinela tawang nggole, Krokot, Kucai, Kucai Muda (Lokio), Kuda daging segar, Kue Koya, Kue sus, Kulit melinjo, Kundur segar, Kunyit, Kura-kura, Langsat, Lantar segar, Lawar babi masakan, Lawar penyu masakan, Lawara jangang masakan, Lawara penjah masakan, Lema/ Rebung asam, Lemon segar, Lemon Squasih, Lemonade, Leunca buah, Lilin bungkus gedi, Lobak, Lokan segar, Lontar segar, Lumai/Leunca segar, Mangga benggala segar, Oramu ninahu ndawa olaho masakan, Otak, Paku hinela wulelenggapaya, Pala daging, Papais, Papeda, Parede baleh masakan, Paria (Pare), Paria Putih kukus, Paria putih segar, Pe-Cay, Peda Ikan Banjar, Pelecing kangkong, Pelepah manuk masakan, Pempek adaan, Pempek kapal selam, Pempek kelesan, Penyu serapah masakan, Pepare ular segar, Pepaya Muda, Pepaya segar, Pucuk lumai/daun leunca segar, Punai daging goreng, Pundut nasi, Purundawa, Purut segar, Putri malu segar, Putri selat, Ragi, Rajungan segar, Rambutan, Rambutan Aceh, Rambutan binjai segar, Rambutan sinyonya, Rawon masakan, Rebon (udang kecil segar), Rebung, Rendang sapi masakan, Rimbang segar, Rujak cingur, Rumput laut, Rusip, RW anjing masakan, Uceng/ bunga melinjo segar, Udang besar segar, Udang galah segar, Udang segar, Ulat sagu segar, Umbut rotan, Usus Sapi, Uwi, Waluh balamak, Woku ubi, Wortel Segar, Wortel kukus, Wortel rebus, Yoghurt.	According to this cluster, 80 grams of food had a high calorie level (above 100 to 500 calories), an average protein and fat content (above 2 to 50 mg), and a carbohydrate content (below 5 mg).

Cluster	List Of Food	Description
2	Abon, Abon haruwan, Ampas Tahu, Angsa, Arwan sirsir, Ayam, Ayam goreng Kentucky sayap, Babi hutan masak rica masakan, Bebek (itik), Bebek alabio daging segar, Bebek daging goreng, Belut goreng, Buaya daging dendeng mentah, Bubur, Bungkil Kacang Tanah, Cassavastick, Coklat Pahit batang, Corned Beef, Cumi-cumi goreng, Daging Babi Gemuk, Daging Babi Kurus, Dendeng belut goreng, Dendeng Daging Sapi, Dendeng mujahir goreng, Domba daging gemuk segar, Empal (daging) Goreng masakan, Empal Goreng, Enting-enting wijen, Gendar goreng, Ham, Itik alabio daging dendeng mentah, Jambal goreng, Kacang atom, Kacang belimbing (kecipir) kering, Kacang goyang, Kacang Kedelai basah, Kaholeo masakan, Kecipir biji, Keju, Keju Kacang Tanah, Kelapa hutan kering, Kerupuk urat, Kluwek, Kripik Tempe Goreng, Kue kelapa, Kwaci, Kwark (Quark), Lamtoro var. gung tanpa kulit, Pala biji, Paniki masak santan masakan, Pencok lele masakan, Pisang Siam goreng, Rebon kering, Rusa daging dendeng mentah, Saga biji tanpa kulit, Santan murni, Sapi daging dendeng mentah, Sapi daging gemuk segar, Sardines dalam kaleng, Sari Kedelai bubuk, Sate Bandeng, Sate pusut masakan, Sie reuboh masakan, Susu bubuk, Tahu telur, Teri balado masakan, Teripang dendeng mentah, Terong + Oncom makanan, Tinoransak masakan, Toge -Tahu makanan, Udang kering, Udang kering mentah, Udang papay mentah, Ulat sagu panggang, Wijen, Worst (sosis daging).	This cluster discovered that the dietary data had enough carbohydrate levels (average of over 1 mg) from 90 grams of food, protein levels (below 60 mg), fat levels (average of below 40 mg), and calorie levels (between 4 and 300 cal).

Three clusters with disparate outcomes are derived from the K-Medoids algorithm's implementation results. Foods with high calorie and carbohydrate content are found in Cluster 0. They have minimal quantities of fat and protein. According to this cluster, food data has a high calorie content (60–720 calories), low protein and fat content (average <50 mg), and high carbohydrate content (>300 mg) per 100 grams of food.Foods having a high calorie content and relatively low protein, fat, and carbohydrate contents are also abundant in cluster 1. According to this cluster, 80 grams of food had a high calorie level (above 100 kcal to 500 kcal), an average protein and fat content (above 2 mg to 50 mg), and a carbohydrate content (below 5 mg).Foods with very high calorie content and relatively low levels of protein, fat, and carbohydrates make up the final cluster, cluster 2. According to this cluster, dietary data had nutritional levels ranging from 4 to 300 calories, less than 60 mg of protein, less than 40 mg of fat, and more than 1 mg of carbs on average from 90 grams of food.

IV. CONCLUSION

The ideal number of clusters (k) in clustering algorithms like K-Means and K-Medoids is determined by the experimental results of the Elbow technique. Plotting the sum of squared errors (WCSS) versus the number of clusters (k) and examining the elbow points on the graph is how this method operates. The elbow point is typically the ideal k value since it shows the k value at which adding more clusters does not considerably and drastically lower the WCSS. The elbow point at cluster k = 3 is displayed by the elbow method on the K-means algorithm with a value of 0.033, while the K-medoids algorithm with the same cluster k has a value of 0.046. The K-Means algorithm's elbow approach yielded a value of 0.033, the smallest k value with the number of clusters k = 3, according to the findings of the comparison between the K-Means and K-Medoids algorithms.

This demonstrates how well the K-Means algorithm clusters foods according to their nutritional worth. The K-Means algorithm's elbow approach yielded a value of 0.033, the smallest k value with the number of clusters k = 3, according to the findings of the comparison between the K-Means and K-Medoids algorithms. This demonstrates how well the K-Means algorithm clusters foods according to their nutritional worth. The K-Medoids algorithm yielded the smallest value of 0.631 in the comparison results between the K-Means and K-Medoids algorithms when the Davies-Bouldin Index method was used, while the K-Means algorithm with the same method yielded the smallest value of 0.643.

V. ACKNOWLEDGMENTS

Sincere gratitude is extended to everyone who has assisted and encouraged our research endeavor. We would especially like to thank Politeknik Harapan Bersama for funding the research through a grant for the Beginner Lecturer Research program.

REFERENCES

[1] A. Pratama, "Noodle Grouping Based On Nutritional Similarity With Hierarchical Cluster Analysis Method," *Anal. Kekuatan Kontruksi Rangka pada Peranc. Belt Conveyor Menggunakan Ansys Work.*, vol. 2, no. 2, pp. 58–71, 2023.

[2] S. Langyan, P. Yadava, F. N. Khan, Z. A. Dar, R. Singh, and A. Kumar, "Sustaining Protein Nutrition Through Plant-Based Foods," *Front. Nutr.*, vol. 8, no. January, 2022, doi: 10.3389/fnut.2021.772573.

[3] P. Alga Vredizon, H. Firmansyah, N. Shafira Salsabila, and W. Eko Nugroho, "Penerapan Algoritma K-Means Untuk Mengelompokkan Makanan Berdasarkan Nilai Nutrisi," ***J. Technol. Informatics,*** vol. 5, no. 2, pp. 108–115, 2024, doi: 10.37802/joti.v5i2.577.

[4] S. S. Gropper, "The Role of Nutrition in Chronic Disease," *Nutrients*, vol. 15, no. 3, pp. 15–17, 2023.

[5] S. S. Nagari and L. Inayati, "Implementation of Clustering Using K-Means Method To Determine Nutritional Status," ***J. Biometrika dan Kependud.***, vol. 9, no. 1, pp. 62–68, 2020, doi: 10.20473/jbk.v9i1.2020.62-68.

[6] A. Hoerunnisa, G. Dwilestari, F. Dikananda, H. Sunana, and D. Pratama, "Komparasi Algoritma K-Means Dan K-Medoids Dalam Analisis Pengelompokan Daerah Rawan Kriminalitas Di Indonesia," ***JATI (Jurnal Mhs. Tek. Inform.***, vol. 8, no. 1, pp. 103–110, 2024, doi: 10.36040/jati.v8i1.8249.

[7] Irwanto, Y. Purwananto, and R. Soelaiman, "Optimasi Kinerja Algoritma Klasterisasi K-Means," ***J. Tek. ITS,*** vol. 1, no. 1, pp. 197–202, 2012.

[8] M. Orisa, "Optimasi Cluster pada Algoritma K-Means," *Pros. SENIATI*, vol. 6, no. 2, pp. 430–437, 2022.

[9] I. W. Septiani, A. C. Fauzan, and M. M. Huda, "Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru," ***J. Sist. Komput. dan Inform.,*** vol. 3, no. 4, p. 556, 2022, doi: 10.30865/json.v3i4.4055.

[10] D. Hartama and I. S. Damanik, "Pengelompokan Algoritma K-Means dan K-Medoid Berdasarkan Lokasi Daerah Rawan Bencana di Indonesia dengan Optimasi Elbow , DBI , dan Silhouette," vol. 6, no. 2, pp. 1151–1158, 2024, doi: 10.47065/bits.v6i2.5851.

[11] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," ***J. Inf. Syst. Informatics,*** vol. 3, no. 3, pp. 424–439, 2021.

[12] B. A. Setiawan and Sulastri, "Perbandingan Clustering Optimalisasi Stok Barang Menggunakan Algoritma K – Means Dan Algoritma K – Medoids," pp. 978–979, 2021, [Online]. Available: https://unisbank.ac.id/ojs/index.php/sendi_u/article/view/8645

[13] P. Vania and B. Nurina Sari, "Perbandingan Metode Elbow dan Silhouette untuk Penentuan Jumlah Klaster yang Optimal pada Clustering Produksi Padi menggunakan Algoritma K-Means," ***J. Ilm. Wahana Pendidik.***, vol. 9, no. 21, pp. 547–558, 2023, [Online]. Available: https://doi.org/10.5281/zenodo.10081332

[14] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," ***Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.***, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.

[15] F. Irhamni, F. Damayanti, B. Khusnul K, and M. A, "Optimalisasi Pengelompokan Kecamatan Berdasarkan Indikator Pendidikan Menggunakan Metode Clustering dan Davies Bouldin Index," *Semin. Nas. dan Teknol. UMJ*, no. 11, pp. 1–5, 2014.

[16] I. T. Umagapi, B. Umaternate, S. Komputer, P. Pasca Sarjana Universitas Handayani, B. Kepegawaian Daerah Kabupaten Pulau Morotai, and B. Riset dan Inovasi, "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa," *Semin. Nas. Sist. Inf. dan Teknol.*, vol. 7, no. 1, pp. 303–308, 2023.

[17] I. F. Fauzi, M. Gito Resmi, and T. I. Hermanto, "Penentuan Jumlah Cluster Optimal Menggunakan Davies Bouldin Index pada Algoritma K-Means untuk Menentukan Kelompok Penyakit," vol. 7, no. 2, pp. 2598–8069, 2023.

[18] A. R. Aruadi, F. Andreas, and B. N. Sari, "Analisis Klaster K-Means pada Data Rata-Rata Konsumsi Kalori dan Protein Menurut Provinsi dengan Metode Davies Bouldin Index," ***J. Pendidik. dan Konseling,*** vol. 4, no. 4, pp. 5652–5661, 2022.