Zeta – Math Journal
Volume 10 No. 1, May 2025, pp. 19-29
E-ISSN: 2579-5864 P-ISSN: 2459-9948
https://doi.org/10.31102/zeta.2025.10.1.19-29

Implementation of the K-Means Method for Beverage Clustering Based on Calorie and Protein

Anggita Eka Rewina[1], Rinci Kembang Hapsari[1*], Chatarina Natassya Putri[1], Gamaliel Virani Fofid Lande[1], Andre Fransisco Aditya[1], Mochamad Tegar Bagas Alamsyah[1]

[1]Prodi Teknik Informatika, Fakultas Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya, Surabaya, Indonesia

**Corresponding Author's Email: rincikembang@itats.ac.id*

ABSTRACT

Recently, the number of coffee shops in big cities in Indonesia has increased. This makes it easier for coffee lovers to enjoy it. With the increasing public awareness of the importance of healthy drinking patterns in preventing diabetes and other diseases, consuming low-calorie drinks has become a prominent trend. This study aims to group the coffee drink menu at Starbucks based on the calorie and protein content of Starbucks drinks. It is grouped into 2 clusters, namely, high and low clusters. In this study, the clustering process of Starbucks drink menu data was carried out by applying the K-Means algorithm. The clustering results can identify members of Cluster 1 and members of Cluster 2. From the tests that have been carried out, it can group the drink menu into 2 clusters based on the amount of protein and calories from Starbucks drinks and help the public choose which drinks are better to consume.

Keyword: *Beverage content; Cluster, Clustering; K-Means,*

Article info:
Submitted: November 21, 2024
Accepted: May 27, 2025

How to cite this article:
Rewina, A. E., Hapsari, R. K., Putri, C. N, Lande, G. V. F., Aditya, A. F., & Alamsyah, M. T. B. (2025). Implementation of the K-Means Method for Beverage Clustering Based on Calorie and Protein. Zeta - Math Journal, 10(1), 19-29.
https://doi.org/10.31102/zeta.2025.10.1.19-29

1. INTRODUCTION

In this modern era, information about the nutritional content of food and beverage products is becoming increasingly important for consumers who care about health (Permana et al., 2017). Consumers pay attention not only to taste and price but also to their drinks' calorie and protein content. The development of innovation in the beverage industry, especially in drinks that contain high calories and protein, has increased significantly (Popkin et al., 2021). As one of the latest innovation drivers in the beverage industry, Starbucks provides coffee and is actively developing a variety of coffee-based and non-coffee drinks (Azriuddin et al., 2020).

The brand's nutritional content is also inseparable from its high calories and protein. Calories and protein play essential roles in human nutrition. Calories are a combination of macronutrients derived from protein, carbohydrates, and fat, while protein is an energy producer of many amino acids (Aryadi et al., 2022).

As the 3rd highest sweetener consumer in the world, Indonesia has a high demand for sweet drinks (Ferretti & Mariani, 2019). The public needs to increase awareness of consuming these brands. Efforts to overcome this problem can utilize the clustering method using the K-Means algorithm.

Previous studies have implemented the K-Mean method to identify essential performance factors that companies must demonstrate to meet consumer satisfaction (Arifin & Agustin, 2023). Classification acts as a process of grouping objects, entities, or information based on characteristics of similarities or differences to facilitate understanding and analysis through the arrangement of more structured categories (Yaumi et al., 2020). The K-Means algorithm is an effective method for grouping data based on similar characteristics. K-Means works by dividing data into several groups (clusters) based on the closest distance to the cluster center (centroid) (Falakhi, 2023).

K-Mean method analysis is helpful in the business world, where business actors must think of ways to continue to survive and increase the scale of their business (Alvisan, 2021). In its application, K-Means can also be used to group the nutritional status of toddlers (Nagari & Inayati, 2020). Consumer segmentation is a crucial step in marketing strategy, where data mining plays an essential role in supporting this grouping process (Amborowati & Winarko, 2014).

The clustering method can cluster several data into clusters based on the level of similarity in one cluster and in another cluster, likewise at a high level of dissimilarity (Aryadi et al., 2022). The use of the K-Means algorithm is considered to be able to run data grouping that refers to several types of data distribution patterns/trends; therefore, the relationship between attributes is found between one another (Negara et al., 2021).

The K-Mean method, used in conjunction with the elbow method, can provide optimal information in beverage clustering, resulting in a deeper understanding of consumer preferences for nutritional content in beverages (Yucha & Safitri, 2021). Optimal clustering using the elbow method can help companies make more targeted decisions in product development and marketing strategies (Yaumi et al., 2020).

The use of the K-Means algorithm is considered to have advantages in implementation and relatively fast processing time (Mursalim et al., 2021). The K-Means algorithm also has advantages in finding ideal clusters but has several shortcomings in determining the centroid point and k value, thus affecting suboptimal performance (Ariasa et al., 2020). To improve the performance of the K-Means algorithm, you can use the elbow method to get the optimal k value (Trianto et al., 2023). The K-Means algorithm can provide valuable information for consumers, the beverage industry, and related parties. Using the K-Means algorithm, drinks can be grouped based on their calorie and protein content so consumers can more easily choose drinks that suit their needs and preferences. Based on previous research, this study aims to obtain the results of grouping types of Starbucks drinks with high calorie and protein content. The K-Means algorithm was used as a grouping method, and the evaluation was done by applying the Elbow method to optimize the search for a good k value.

2. LITERATURE REVIEW

2.1. Data Mining

Data mining is exploring and analyzing big data to find patterns, correlations, trends, and helpful information that can be used for decision-making. Data mining combines techniques from various disciplines, such as statistics, machine learning, and artificial intelligence, to analyze large and complex datasets (Zulfa et al., 2021). Data mining is a potent tool for uncovering hidden insights in data and supporting data-driven decision-making. With the advancement of technology and the increasing volume of available data, the ability to process and analyze data is becoming increasingly important. For organizations that can use it properly, data mining is not just about finding patterns in numbers but also opening the door to innovation and competitive advantage in an increasingly dynamic market.

2.2. Pre-processing Data

Data preprocessing is the first step before performing analysis or applying machine learning algorithms. This process includes a series of steps to clean, transform, and prepare data so that it is ready for use in more sophisticated analysis or prediction models. Without proper data preprocessing, the model built can produce poor or even misleading results.

Data transformation is the process of changing raw data into a form that is more suitable or easier to analyze or process. It aims to improve data quality, reduce noise, or make the data more suitable for the model used. This transformation can take the form of changes to the data scale, data grouping, creating new features, or changing the data type to make it more compatible with a particular analysis or algorithm.

One of the data transformation processes is data normalization; normalization is a method used to process data by creating several variables with the same range of values (Harmain et al., 2021); (Hapsari et al., 2023). This dramatically affects the results of processing because if there is a vast difference in the range of values between the other attributes, it will cause the attribute that has a larger value to have a powerful influence on the attribute that has a small value (Aziz et al., 2021)

There are also methods for normalization, namely Z-score Normalization, Min-Max normalization, and Decimal Scaling Normalization (Ali, 2022). The min-max normalization method standardizes the data on the samples taken so that each value in the calorie and protein variables has the same scale (Permana & Salisah, 2022). Min-max normalization is a technique by performing linear calculations on the original data to create a value proportion between other attributes (Mangku Negara et al., 2021)

2.3. Clustering

Clustering, or grouping, is a data analysis technique that aims to group objects or data points into groups (clusters) based on the similarity of their characteristics (Mulyana et al., 2024). Clustering is a handy tool in various disciplines, allowing the identification of hidden structures, patterns, and relationships in complex data sets (Nagari & Inayati, 2020).

In clustering, data is grouped into clusters based on maximizing intra-cluster similarity (data in the same cluster have high similarity) and minimizing inter-cluster similarity (data between clusters have significant differences). In other words (Ashari et al., 2019), objects in the same cluster should be very similar, while objects in different clusters should be very different (Revathi et al., 2021). Clustering allows a deeper understanding of the structure of the data and helps in better decision-making.

Clustering is the process of grouping a set of data into similar groups. This process helps in understanding and finding natural group patterns in a dataset. Clustering algorithms have several categories based on features and mathematical modeling. One is centroid-based clustering using the K-Means algorithm (Uykan, 2023).

Clustering algorithms vary in their approaches and underlying assumptions, which affect the types of clusters found and their effectiveness on certain types of data. Hierarchical clustering methods build a hierarchical cluster hierarchy in stages while partitioning methods divide data into a predetermined number of clusters. Clustering is crucial in various real-world applications, including customer segmentation, document clustering, image analysis, and bioinformatics. In customer segmentation, clustering is used to identify distinct customer groups based on their purchasing behavior, demographics, or preferences.

The success of clustering depends heavily on selecting relevant features, appropriate distance measures, and the right clustering algorithm (Falakhi, 2023; Yaumi et al., 2020). Evaluation of clustering results is also essential to ensure that the resulting clusters are meaningful and by the objectives of the analysis.

3. METHODE

In this research, four processes were carried out: dataset, data preprocessing, clustering, and system performance measurement.

3.1 Dataset

The dataset used in this research is secondary data sourced from the official Starbucks website, where the data was obtained from the nutritional content information at Starbucks. The data amounts to 1380 based on the type of drink at Starbucks. In the dataset, there are variables energy (kcal), fat (grams), of which saturated (grams), carbohydrates (grams), of which sugar (grams), fiber (grams), protein (grams), salt (grams), and caffeine (grams).

3.2 Preprocessing Data

At this stage, preprocessing performs two processes, namely sampling and normalization. Sampling is done using a random technique with the provisions of cup size and taking two attributes: calories and protein. In the second process, normalization is carried out. In this study, data normalization uses the Min-max Normalization method, with formula (1):

$$X_{norm} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \left(Max_{baru} - Min_{baru} \right) + Min_{baru} \quad (1)$$

Where :

- X is the original value.
- X_{min} is the minimum value of the data.
- X_{max} is the maximum value of the data.
- Max_{new} is the new maximum value (e.g. 1)
- Min_{new} is the new minimum value (e.g. 0)

3.3 K-Means

The K-Means algorithm is a clustering technique built using the partition principle. This algorithm divides objects with the same level of similarity into the same cluster, while objects that do not have a high level of difference are placed in different clusters (Abid et al., 2022) (Bangkalang et al., 2023).

The following is the process flow in the K-Means algorithm, according to (Mursalim et al., 2021):

a. Prepare the dataset
b. Determine the number of clusters (k)
c. Initialize the center point (centroid), which is determined randomly
d. Calculate the distance of each data in the dataset to the centroid using the formula (2):

$$d_{ij} = \sqrt{(x_i - x_{cj})^2 + (y_i - y_{cj})^2} \quad (2)$$

Where,

d_{ij} is the distance of the i-th data with the j-th centroid
x_i is the attribute/parameter value x of the i-th data
y_i is the attribute/parameter value y of the i-th data
x_{cj} is the attribute/parameter value x of the j-th centroid
y_{cj} is the attribute/parameter value y of the j-th centroid

e. Determine the group of each data into the nearest centroid
f. Determine the new centroid value by calculating the average value of the cluster members
g. Repeat the d-th process, if there is still data (cluster members) that move clusters or if there is a change in the value of the centroid.

3.4 Performance Measurement

Before implementing K-Means, the first thing to do is to get information on the number of the best k clusters using the elbow method. The purpose of this method is to determine a small k value that also has a low within value. The selection of the optimal number of clusters is done by observing the Sum Squared Error (SSE) value, which is the comparison of the calculation of the sum of the squares of the distance between the centroid and the cluster members between the number of clusters that will form an elbow at a point. The SSE value can be calculated using equation (2).

$$SSE = \sum_{K=1}^{K} \sum_{x_i \in S_K} \|x_i - C_k\|^2 \quad (3)$$

Where,

K is number of clusters
x_i is attribute value/parameter of the i-th data
C_k is attribute value/parameter of the k-th centroid

The more clusters k, the smaller the SSE value becomes. Where are The steps used in the elbow method as follows:

a. Provide/determine the initial K value
b. Add/enlarge the K value
c. Calculate the SSE value from the K value that has decreased drastically
d. Determine the K value in the form of an elbow

4. RESUT AND DISCUSSION

The dataset used in this study is the nutritional information data contained in Starbucks drinks. It consists of 1380 data with 9 variables. From the 1380 data, 99 popular drink menu data were taken at Starbucks, and feature selection was carried out to determine the 2 attributes used, namely the Calories and Protein attributes. The data used are shown in Table 1.

The data above is sample data selected based on the calorie and protein content of the Starbucks drink menu. This data determines whether the drink has a high or low content.

Table 1. Calorie and Protein Content Dataset in Beverages

Product	Calories	Protein
Americano	16	1
Caffe (C) Latte (L)- semi skimmed milk	174	13
CL - whole milk	229	13
CL - skimmed milk	133	13
CL - almond drink	96	3
CL - soya drink	154	12
CL - oat drink	202	4
CL - coconut drink	161	4
Latte Macchiato - semi skimmed milk	174	13
Latte Macchiato - whole milk	229	13
Latte Macchiato - skimmed milk	133	13
Latte Macchiato - coconut drink	161	4
Cappuccino (Cpc) - semi skimmed milk	139	10
Cpc - whole milk	181	10
Cpc - skimmed milk	106	10
Cpc - almond drink	77	2
Cpc - soya drink	123	10
Cpc - oat drink	160	4
Cpc- coconut drink	128	4
Misto - semi skimmed milk	141	10
Misto (M)- whole milk	181	10
M - skimmed milk	111	10
M - almond drink	84	3
M - soya drink	126	10
M - oat drink	161	4
M - coconut drink	131	4
Caramel (Crm) Macchiato (Mch)- whole milk	266	11
Crm Mch - skimmed milk	184	11
Crm Mch - almond drink	151	3
Crm Mch - soya drink	201	11
Crm Mch - oat drink	243	4
Crm Mch - coconut drink	207	4
Mocha - semi skimmed milk	317	13
Mocha - whole milk	353	13
Mocha - skimmed milk	290	13
Mocha - almond drink	265	6
Mch - soya drink	304	13
Mch - oat drink	335	8
Mocha - coconut drink	308	8
White (W) Mch - semi skimmed milk	364	12
W Mch - whole milk	400	12
W Mch - skimmed milk	337	12
W Mch - almond drink	312	5
W Mch - soya drink	351	11
W Mch - oat drink	382	6
W Mch - coconut drink	355	6

Product	Calories	Protein
Cold (C) Brew (B)	47	12
C B L - semi skimmed milk	131	14
C B L - whole milk	167	14
C B L - skimmed milk	105	14
C B L - almond drink	80	8
C B L - soya drink	118	14
C B L - oat drink	149	9
C B L - coconut drink	122	9
Nitro (N) C B	81	21
N L	221	16
N Cpc	92	18
Iced (I) Americano (A)	16	1
I L - semi skimmed milk	121	9
I L - whole milk	157	9
I L - skimmed milk	93	9
I L - almond drink	68	2
I L - soya drink	107	9
I L - oat drink	139	3
I L - coconut drink	111	3
Classic I Cappuccino - semi skimmed milk	121	9
Classic I Cappuccino - whole milk	157	9
Classic I Cappuccino - skimmed milk	93	9
Classic I Cappuccino - almond drink	68	2
Classic I Cappuccino - soya drink	107	9
Classic I Cappuccino - oat drink	139	3
Classic I Cappuccino - coconut drink	111	3
I L Macchiato - semi skimmed milk	121	9
I L Macchiato - whole milk	157	9
I L Macchiato - skimmed milk	93	9
I L Macchiato - almond drink	68	2
I L Macchiato - soya drink	107	9
I L Macchiato - oat drink	139	3
I L Macchiato - coconut drink	111	3
I Caramel Mch - semi skimmed milk	188	9
I Caramel Mch - whole milk	224	9
I Caramel Mch - skimmed milk	160	9
I Caramel Mch - almond drink	135	2
I Caramel Mch - soya drink	174	9
I Caramel Mch - oat drink	206	3
I Caramel Mch - coconut drink	178	3
Iced Cappuccino with Cold Foam	77	5
I Mocha (Mo) - semi skimmed milk	363	14
I Mo - whole milk	399	13
I Mo - skimmed milk	335	14
I Mo - almond drink	310	7
I Mo - soya drink	349	13
I Mo - oat drink	381	8
I Mo - coconut drink	353	8
I White (W) Mo - semi skimmed milk	410	12
I W Mo - whole milk	446	12
I W Mo - skimmed milk	382	12
Coffee Frappuccino® - semi skimmed milk	280	5
Coffee Frappuccino® - whole milk	295	5

The next step is to determine the most appropriate number of k clusters before running the clustering process. This study uses Jupyter, a tool for the Python programming language, to determine the number of clusters. The results are as follows.

```
inert = [ ]
k_range = range(1,49)
for k in k_range:
    km = K_Means(n_klaster=k).fit(x_train)
    inert.append(km.inertia_)
```

The output of the program above produces a graph/diagram that shows the most ideal number of K clusters, therefore the next stage can be carried out, namely clustering using the K-Means algorithm.

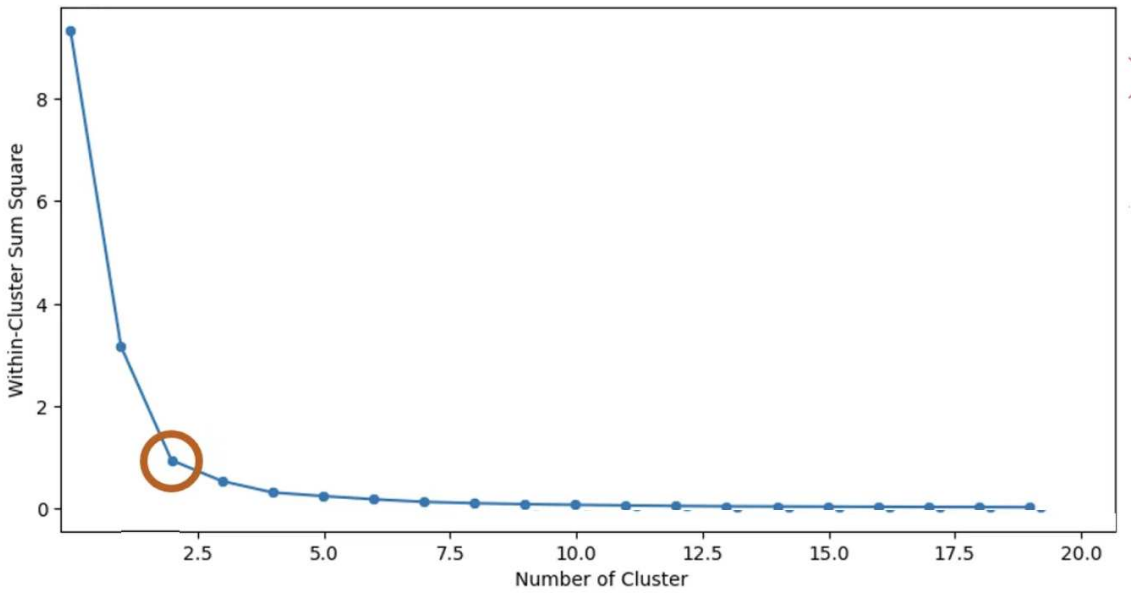

Figure 2. Elbow method graph

Based on Figure 2, it can be seen that the number of K clusters produced based on the results of the elbow method calculation with the Python programming language is 2 clusters. Based on the results obtained, the average data on protein and calorie consumption will be clustered into 2 clusters, consisting of low and high clusters or K = 2. The clustering of two clusters is done so that it is known which drink values are in the low cluster and high cluster. The next stage is the clustering process using K-Means. Figure 3 shows a model of the K-Means clustering results produced with updated centroids.

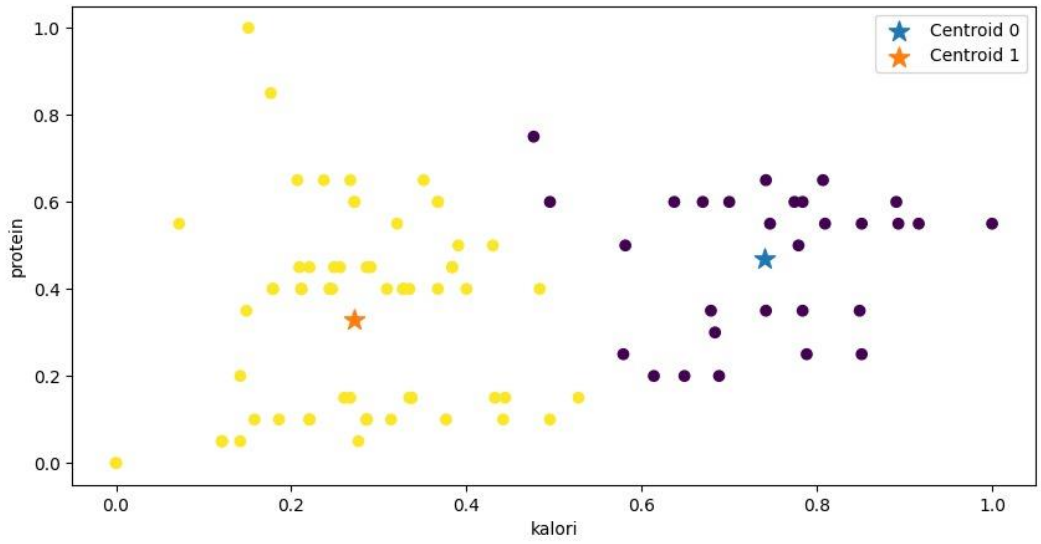

Figure 3. Results of the K-Means algorithm with new centroid results.

Based on the last centroid value, which is the result of the previous process, 2 grouping clusters are generated based on the levels of average protein and calorie data. Figure 3 visualizes the cluster results formed from the K-Means algorithm.

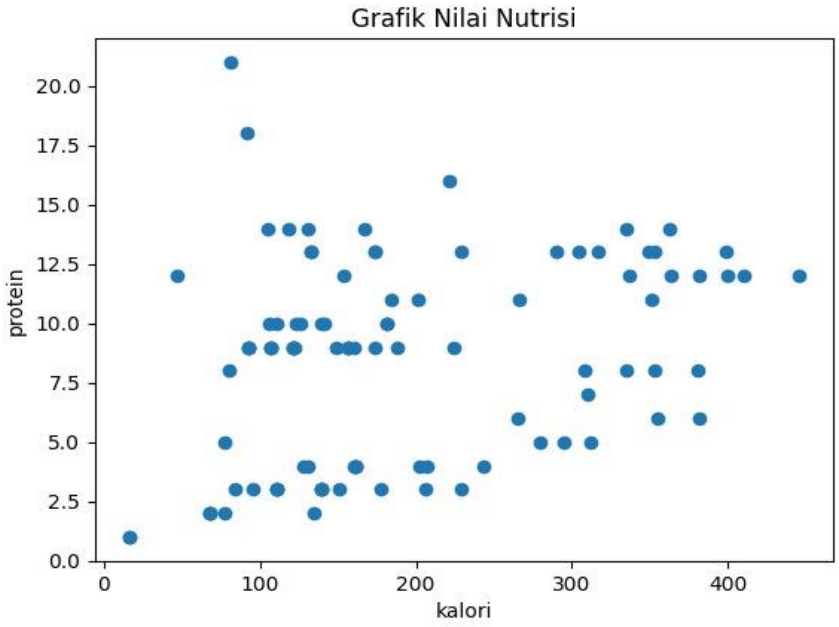

Figure 4 visualization of K-Means cluster results

The results of cluster visualization are analyzed usingition of the cluster distribution data points in the KMeans cluster visualization results. If the distribution data points are getting closer to the left, with a value of 0, it can be concluded that the cluster is a group of drinks with a low-calorie level. Conversely, the calorie level is high if the distribution data points get closer to the right.

5. CONCLUSION

Referring to the research results that have been produced, namely to group Starbucks drinks based on protein and calorie nutritional content, it can be concluded that the K-Means algorithm forms a value of K = 2 clusters. So, the clustering of Starbucks drinks refers to the protein and calorie nutritional content divided into two categories, namely low and high. The results of the K-Means algorithm produced 2 clusters that refer to the protein and calorie nutritional content, namely cluster 1 and cluster 2.

REFERENCE

Ali, P. J. M. (2022). Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor with Different Similarity Measurements. Aro-the Scientific Journal of Koya University, 10(1), 85–91. https://doi.org/10.14500/aro.10955

Alvisan, F. K. (2021). Clustering Minimarket Untuk Menentukan Jumlah Kebutuhan Pembelian Menggunakan Metode K-Means. Nusantara of Engineering (NOE), 4(2), 160. https://doi.org/10.29407/noe.v4i2.16784

Amborowati, A., & Winarko, E. (2014). Review Pemanfaatan Teknik Data Mining Dalam Segmentasi Konsumen. https://ejournal.gunadarma.ac.id/index.php/kommit/article/ download/1009/869

Ariasa, K., Gede, I., Gunadi, A., & Made Candiasa, I. (2020). Optimasi Algoritma Klaster Dinamis Pada K-Means dalam Pengelompokan Kinerja Akademik Mahasiswa (Studi Kasus: Universitas Pendidikan Ganesha). Janapati, 9(2), 181–194.

Arifin, Moh. M., & Agustin, M. (2023). Analisis Strategi Pengembangan Usaha Menggunakan SWOT dan Business Model Canvas Pada Produk Air Minum Dalam Kemasan PT. Abutama. Jurnal KaLIBRASI - Karya Lintas Ilmu Bidang Rekayasa Arsitektur Sipil Industri, 6(1), 1. https://doi.org/10.37721/kalibrasi.v6i1.1072

Aryadi, A. R., Andreas, F., & Sari, B. N. (2022). Analisis Klaster K-Means pada Data Rata-Rata Konsumsi Kalori dan Protein Menurut Provinsi dengan Metode Davies Bouldin Index. Jurnal Pendidikan Dan Konseling, 4(4), 5652–5661.

Ashari, B. S., Otniel, S. C., & Rianto. (2019). Perbandingan Kinerja K-Means Dengan DSCAN Untuk Metode Clustering Data Penjualan Online Retail. Jurnal Siliwangi, 5(2), 72–77. http://jurnal.unsil.ac.id/index.php/jssainstek/article/view/1283

Aziz, A. R., Warsito, B., & Prahutama, A. (2021). Pengaruh Transformasi Data Pada Metode Learning Vector Quantization Terhadap Akurasi Klasifikasi Diagnosis Penyakit Jantung. Jurnal Gaussian, 10(1), 21–30. https://doi.org/10.14710/j.gauss.v10i1.30933

Azriuddin, M., Kee, D. M. H., Hafizzudin, M., Fitri, M., Zakwan, M. A., AlSanousi, D., Kelpia, A., & Kurniawan, O. (2020). Becoming an International Brand: A Case Study of Starbucks. Journal of The Community Development in Asia, 3(1), 33–43. https://doi.org/10.32535/jcda.v3i1.706

Falakhi, A. (2023). Pengolahan Data Pelanggan Dengan Tenik Clustering K-Means Di Aplikasi Weka. Journal Computer Science and Informatic Systems J-Cosys, 3(2), 54. https://doi.org/10.53514/jco.v3i2.394

Ferretti, F., & Mariani, M. (2019). Sugar-sweetened beverage affordability and the prevalence of overweight and obesity in a cross section of countries. Globalization and Health, 15(1), 1–14. https://doi.org/10.1186/s12992-019-0474-x

Hapsari, R. K., Salim, A. H., Meilani, B. D., Indriyani, T., & Rachman, A. (2023). Comparison of the Normalization Method of Data in Classifying Brain Tumors with the k-NN Algorithm. The 2nd International Conference on Neural Networks and Machine Learning, 1, 21–29. https://doi.org/10.2991/978-94-6463-174-6_3

Nagari, S. S., & Inayati, L. (2020). Implementation of Clustering using K-Means Method To Determine Nutritional Status. Jurnal Biometrika Dan Kependudukan, 9(1), 62. https://doi.org/10.20473/jbk.v9i1.2020.62-68

Negara, M., Setyawan, I., Purwono, P., & Ashari, I. A. (2021). Analisa Cluster Data Transaksi Penjualan Minimarket Selama Pandemi Covid-19 dengan Algoritma K-means. JOINTECS (Journal of Information Technology and Computer Science), 6(3), 153. https://doi.org/10.31328/jointecs.v6i3.2693

Mulyana, A., Hermawan, Y., & Saputri, J. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Rekomendasi Pilihan Program Studi Pada Mahasiswa Baru. KERNEL : Jurnal Riset Inovasi Bidang Informatika Dan Pendidikan Informatika, 5(1), 60–72. https://doi.org/10.31284/j.kernel.2024.v5i1.7624

Mursalim, M., Purwanto, P., & Soeleman, M. A. (2021). Penentuan Centroid Awal Pada Algoritma K-Means Dengan Dynamic Artificial Chromosomes Genetic Algorithm Untuk Tuberculosis Dataset. Techno.Com, 20(1), 97–108. https://doi.org/10.33633/tc.v20i1.4230

Permana, A. W., Widayanti, S. M., Prabawati, S., & Setyabudi, D. A. (2017). Sifat Antioksidan Bubuk Kulit Buah Manggis (Garcinia Mangostana L.) Instan Dan Aplikasinya Untuk Minuman Fungsional. Jurnal Penelitian Pascapanen Pertanian, 9(2), 88. https://doi.org/10.21082/jpasca.v9n2.2012.88-95

Permana, I., & Salisah, F. N. S. (2022). Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation. Indonesian Journal of Informatic Research and Software Engineering (IJIRSE), 2(1), 67–72. https://doi.org/10.57152/ijirse.v2i1.311

Pokpin, B. M., & Ng, S. W. (2021). Sugar-sweetened beverage taxes : Lessons to date and the future of taxation. PLOS Medicine, Januari, 1–6. https://doi.org/10.1371/journal.pmed.1003412

Revathi, V., Ghutugade, K. B., Vannapuram, R., & Prasanna, B. P. S. (2021). K-Means Algorithm for Clustering of Learners Performance Levels Using Machine Learning Techniques. Revue d Intelligence Artificielle, 35(1), 99. https://doi.org/10.18280/ria.350112

Trianto, R. B., Nugroho, A. S., & Supriyadi, E. (2023). Klasterisasi Menggunakan Algoritma K-Means dan Elbow pada Opini Masyarakat Tentang Kebijakan Sekolah Luring Tahun 2022. INOVTEK Polbeng - Seri Informatika, 8(1), 1. https://doi.org/10.35314/isi.v8i1.2756

Uykan, Z. (2023). Fusion of Centroid-Based Clustering With Graph Clustering: An Expectation-Maximization-Based Hybrid Clustering. IEEE Transactions on Neural Networks and Learning Systems, 34(8), 4068–4082. https://doi.org/10.1109/TNNLS.2021.3121224

Yaumi, Ach. S., Zulfiqkar, Z., & Nugroho, A. (2020). Klasterisasi Karakter Konsumen Terhadap Kecenderungan Pemilihan Produk Menggunakan K-Means. JOINTECS (Journal of Information Technology and Computer Science), 5(3), 195. https://doi.org/10.31328/jointecs.v5i3.1523

Yucha, N., & Safitri, S. N. (2021). Pengaruh Bauran Pemasaran, Suasana Kafe dan Minat Beli Terhadap Keputusan Pembelian di Intano Coffe Shop and Roastery Krian. IQTISHADequity Jurnal MANAJEMEN, 3(1), 180. https://doi.org/10.51804/iej.v3i1.927

Zulfa, N., Auliya, R. I., & Zaenal, A. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. Jurnal Teknologi Dan Sistem Informasi (JTSI), 2(2), 100. http://jim.teknokrat.ac.id/index.php/JTSI