JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 1, Desember 2022, hlm. 109 - 116 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i1.1855
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

WEB-BASED CLUSTER OPTIMIZATION USING K-MEDOIDS AND DAVIES BOULDIN INDEX

Ryan Christian[1], Deny Jollyta[1*]
[1]Informatics, Institut Bisnis dan Teknologi Pelita Indonesia
email: [*]deny.jollyta@lecturer.pelitaindonesia.ac.id[2]

Abstract: Clustering data has always been a fascinating subject to research numerous perspectives. A variety of knowledge is produced by the calculating process utilizing various algorithms. The genesis of cluster optimization is based on differences of opinion about the cluster's results. In general, cluster and optimization findings are generated using software such as Matlab, RapidMiner, and programming languages like Python. Users, however, have not been satisfied with the results so far. The various outcomes are the primary motivations for continuing to create and develop applications. The goal of this research is to create an application that can evaluate cluster data using the K-Medoids method, which can then be further optimized using the Davies Bouldin Index (DBI). Because the target application is students and lecturers who use it in learning and observers of the cluster field, the application can indeed be accessible through a browser to make it easier to use. For ease of using it, the program is available on both desktop and mobile platforms. Through separately created applications, it is intended that this research will give an alternative to clustering and optimization.

Keywords: application, cluster, dbi, k-medoids, optimization

Abstrak: Clusterisasi data selalu menjadi topik yang menarik untuk dikembangkan dari berbagai sisi. Proses perhitungannya yang menggunakan berbagai algoritma menghasilkan *knowledge* yang beragam. Perbedaan pendapat terhasil hasil cluster menjadi dasar munculnya optimalisasi *cluster*. Umumnya hasil *cluster* dan optimalisasi diperoleh dari pengolahan menggunakan aplikasi yakni *Matlab*, *RapidMiner*, dan bahasa pemrograman seperti *Pyhton*. Namun demikian hasil yang muncul belum mampu memuaskan pengguna. Hasil yang berbeda menjadi alasan utama pembuatan maupun pengembangan aplikasi masih terus dilakukan. Penelitian ini bertujuan untuk membangun sebuah aplikasi yang dapat memproses data *cluster* menggunakan algoritma *K-Medoids* untuk selanjutnya dioptimalisasi dengan *Davies Bouldin Index* (DBI). Untuk memudahkan penggunaan, aplikasi dapat diakses pada browser karena target aplikasi adalah mahasiswa dan dosen yang menggunakan pada pembelajaran serta pemerhati bidang *cluster*. Aplikasi dirancang pada platform *desktop* dan *mobile* demi memudahkan pengaksesan. Diharapkan, penelitian ini memberikan alternatif dalam proses clusterisasi dan optimalisasi melalui aplikasi yang dirancang mandiri.

Kata kunci: aplikasi; *cluster*; dbi; *k-medoids*; optimalisasi

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 1, Desember 2022, hlm. 109 - 116 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i1.1855
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

The grouping outcomes produced by cluster analysis are always problematic. [1]. The outcome is significantly influenced by the employment of different algorithms, measurement techniques, and cluster optimization strategies. Consequently, no particular size is used to decide whether to accept cluster results. [2]. Knowledge or information that is deemed improper frequently causes problems in cluster analysis. This is affected by a number of factors, including the clustering met the results of the 2016 E-Government mapping using the SPSS application hods utilized, measurement techniques, cluster optimization, and the applications used for data processing.

Applications for cluster processing are expanding constantly. To assist interested parties in processing data and producing the needed information, numerous self-designed programs have also been developed in addition to common tools like Matlab, RapidMiner, and computer languages like Python. For one calculation only, some customers prefer a dedicated program rather than having to navigate menus. Developers of applications now have the opportunity to produce applications as needed. This work aims to provide a cluster analysis application for users that are interested in the Manhattan distance formula, the K-Medoids algorithm, and optimized using DBI. Users can access the created application straight through a browser to make things simpler.

K-Medoids and DBI have been used in a variety of cluster analysis investigations, both separately and together. According to research [3], the YouTube channels in Indonesia with the best cluster for their qualities are those with a grade A, a gold rating, and an annual revenue of between $5-$10 million. When optimizing the number of central points with the K-Means method, DBI performs exceptionally well, according to study [4], For the purpose of categorizing Central Java's districts and citiesSelain itu, [5] also uses K-Means and DBI.

In a study [6], research on applications of cluster analysis with Manhattan and DBI is presented. Web-based applications and PHP programming are employed by the Covid-19 task force as the application of this research. To create the best cluster traffic management system, DBI is calculated using a Matlab application in [7], the 2016 E-Government mapping results were used with the SPSS tool in a study carried out by [8].

It is still possible to create cluster analysis apps based on user demands as the foundation for this research, according to the numerous references provided. The application is designed to be used on both desktop and mobile platforms, and it is anticipated to make it easier for users to create cluster analyses using K-Medoids and DBI.

METHOD

In order to accomplish the research aims, a systematic and guided framework of thought was used to conduct the study. The framework in question is shown in Image 1.

The research reviewing the literature on clusters, K-Medoids algorithms, DBI, web programming methodologies, and other supporting materials. The objects known as K-Medoids are thought to represent both cluster centers and clusters. By measuring the similarity distance between medoids and non-medoids, the K-Medoids method creates a cluster [9].

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 1, Desember 2022, hlm. 109 - 116 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i1.1855
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

The following equation applies the Manhattan distance measurement technique to determine the exact distance between two objects' coordinates [10].

$$d(x,y) = \sum_{i-1}^{n} |x_i - y_i| \qquad (1)$$

The following DBI formula is used for the optimization process: [11]:

$$DBI = \frac{1}{K}\sum_{i=1}^{k} max_{i \neq j}(R_{i,j}) \qquad (2)$$

The next step is to analyze the demands of the application. Prior to beginning the design phase, an analysis is performed to determine the application's user requirements. Design is necessary so that the application's appearance can reflect users' demands generally.

The following action is to develop a program for the application in accordance with the design. It is crucial to clearly define the user's application needs. Input of k test counts, algorithm computations, and DBI to dynamic graphic displays are just a few examples of how menus emerge to make operation with interactive and guiding languages easier, include an error notice as a user confirmation of the application. Additionally, data entry is done to evaluate how well the application works. The data used is 510 data points worth of drug user data from UCI machine learning. The program is evaluated once again using the Black Box to ascertain its usability from the user's perspective if successful in creating an ideal cluster. [12]. If the Black Box test is unsuccessful, it is required to assess the application requirements analysis to make sure that it satisfies the user's expectations.

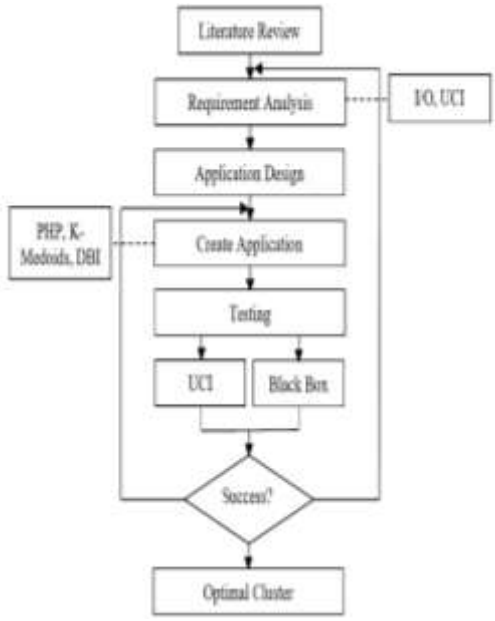

Image 1. Research Stages

RESULT AND DISCUSSION

Application Testing Using Data

The calculation of drug data is displayed in this part straight from the submitted application. The dashboard appears when the application first launches by Image 2.

Image 2. Dashboard Display

K-Medoids are briefly described in Image 2. The choices for the Data Field, Data Value, and Calculation are on the dashboard.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 1, Desember 2022, hlm. 109 - 116 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i1.1855
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Image 3. Data Field Display

Data Field is a feature that allows users to enter field names in accordance with their data demands. The fields in the drug data used are Drug Name, Rating, Effectiveness, and Side Effects. Image 3's data fields can be deleted if the user wants to replace the data, and then they can get the excel template from the Download Template Excel option.

Drug data can be entered by the user directly on the application screen using the Data Value menu, as shown in Image 4. The drug data can also be entered into templates using the Download Template Excel option, which can subsequently be imported using the Import from Excel Template menu.

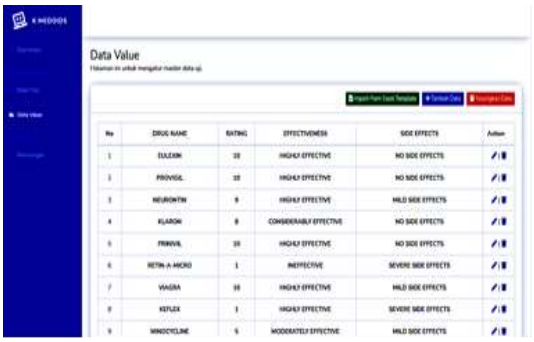

Image 4. Data Value Display

On the other hand, an error message will show up as shown in Image 5 if the data value filling in is not complete according to the designated field.

Image 5. Error Alert

The idea of Image 5 is to lessen user data entry errors. In order to acquire the best cluster results, the error notice advises the user to be extra diligent and finish the data.

The Image 6 is the menu display for calculations. Using the K-Medoids, Manhattan, and DBI algorithms, this menu processes drug data or other data.

Image 6. Calculation

The user selects the number of k tests using the menu in Image 6 before the calculation starts. Ten k tests, with k values ranging from 2 to 11, are used in this study. The process button on the calculation menu has coding that is sequentially organized structure and simultaneously performs 4 tasks, namely: 1). For clustering purposes, transformation is the process of turning character data into integers; 2). Normalization is the process of equating data spacing; 3). Clustering is a calculation to produce groupings using equation (1); 4). Optimization is finding optimal clusters from a number of k tests that have been obtained from grouping using equation (2).

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 1, Desember 2022, hlm. 109 - 116 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i1.1855
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

The results of all processes are the DBI values for each k test as shown in Table 1.

Table 1. DBI Values for Each k Test

K	DBI Values
2	0.327196
3	0.372136
4	0.321875
5	0.284278
6	0.218056
7	0.257786
8	0.143115
9	**0.1337**
10	0.214803
11	0.163348

According to Table 1, the lowest DBI value, or 0.1337, is found when k test = 9. Therefore, the k test of 9 was chosen as the ideal cluster. Image 7 shows the DBI chart in detail.

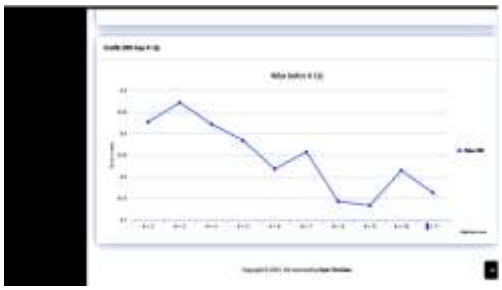

Image 7. DBI Graph

K test = 9 contains cluster members which are formed from calculations of the K-Medoids algorithm with Manhattan distance calculations. The following list represents the members of the ideal k test's cluster:

Table 2. Cluster Members for k Test = 9

No	Drug Name	Rating	Effectiveness	Side Effects	Cluster
15	Motrin	8	Moderately Effective	No Side Effects	1
191	Synthroid	8	Moderately Effective	Mild Side Effects	1
240	Hydrochloro-thiazide	9	Moderately Effective	No Side Effects	1
…					
3	Neurontin	9	Highly Effective	Mild Side Effects	2
7	Viagra	10	Highly Effective	Mild Side Effects	2
…					
19	Prozac	9	Highly Effective	Mild Side Effects	2
24	Tylenol	9	Highly Effective	Mild Side Effects	2
…					
496	Levoxyl	10	Highly Effective	No Side Effects	7

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 1, Desember 2022, hlm. 109 - 116 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i1.1855
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Table 3. Black Box Testing Results

Scenario	Case	Expected Result	Testing Result	Conclusion
Input training data	Input training data on value menu	Data successfully entered	Suitable	Success
	Input training data on excel templete	Data successfully entered	Suitable	Success
Data	300	The application functions properly.	Suitable	Success
	500	The application functions properly.	Suitable	Success
	>1000	The application functions properly.	Suitable	Success
K test total to determine the DBI result	<=5	Quick DBI result display	Suitable	Success
	<=10	Quick DBI result display	Suitable	Success
	<=15	Quick DBI result display	Suitable	Success
	<=20	Quick DBI result display	Suitable	Success
	<=25	Quick DBI result display	Slow DBI result display	Not as effective
	>25	Quick DBI result display	No DBI resukt display	Unsuccessfull
Incomplete data filling	Incomplete data filling	Error Alert	Error Alert	Success

Based on Table 2, the drug data are grouped according to the closeness of each data. However, there are clusters that do not have members, namely clusters 3, 8 and 9.

Black Box Testing

Black Box testing is conducted after this cluster optimization program. Table 3 includes the results of the Black Box test.

The outcomes of the Black Box test were successful, as shown in Table 3. Testing data on k tests over 25 is the application's limitation. Testing on k tests greater than 20 and less than 25 can display DBI findings, but it takes longer. When k tests exceed 25, the application

cannot function properly. The features of the application server running this application may have an impact on this.

CONCLUSION

The K-Medoids method, Manhattan measurement, and web-based DBI cluster optimization were successfully used in this research to build a cluster analysis application. The application displays accurate calculation results and assumptions based on the tests that were done. Data on drug use obtained from UCI demonstrates that the application can execute up to 25 k tests for any volume of data. The smaller the k test that may be employed, the more data there are. This is a drawback of the application as it was intended. The application runs more slowly the more k tests there are.

The application's prepared features all perform flawlessly, though. The KMDBI Apps app, which is accessible via the kmdbi.ryanchristiann.com browser, may be utilized by users without difficulty, according to the Black Box results. The program is set up so that cluster results are not kept in the database, allowing users to run tests on multiple sets of data. Instead, an excel export menu is offered. Automatically, the application chooses a random center point. In order to process cluster data utilizing the Manhattan measurement, DBI, and K-Medoids method, it is envisaged that users will have access to this program.

BIBLIOGRAPHY

[1] D. Jollyta, S. Efendi, M. Zarlis, and H. Mawengkang, "A Combinatorial Optimization Approach to Determining Optimal Data in Cluster," in *International Conference on Artificial Intelligence and MechatronicsSystem (AIMS)*, 2021, pp. 1–5.

[2] J. Ponce and A. Karahoca, *Data Mining and Knowledge Discovery in Real Life Applications*. 2009.

[3] A. Badruttamam, S. Sudarno, and D. A. I. Maruddani, "Penerapan Analisis Klaster K-Modes Dengan Validasi Davies Boundin Index Dalam Menentukan Katakteristik Kanal Youtube Di Indonesia," *J. Gaussian*, vol. 9, no. 3, pp. 263–272, 2020, doi: 10.14710/j.gauss.v9i3.28907.

[4] B. Jumadi Dehotman Sitompul, O. Salim Sitompul, and P. Sihombing, "Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1235, no. 1, 2019, doi: 10.1088/1742-6596/1235/1/012015.

[5] R. D. Kusumah, B. Warsito, and M. A. Mukid, "Perbandingan Metode K-Means Dan Self Organizing Map (Studi Kasus: Pengelompokan Kabupaten/Kota Di Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia 2015)," *J. Gaussian*, vol. 6, no. 3, pp. 429–437, 2017, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian.

[6] W. Gie and D. Jollyta, "Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster

Menggunakan Davies Bouldin Index : Status Covid-19 Wilayah Riau," *Pros. Semin. Nas. Ris. Dan Inf. Sci. 2020*, vol. 2, no. April, pp. 187–191, 2020.

[7] S. Nawrin, M. Rahatur, and S. Akhter, "Exploreing K-Means with Internal Validity Indexes for Data Clustering in Traffic Management System," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3, 2017, doi: 10.14569/ijacsa.2017.080337.

[8] I. D. Apriliyaningsih and D. Istiawan, "Penerapan Seleksi Atribut Berdasarkan Koefisien Variansi dan Korelasi untuk Inisialisasi Pusat Awal Klaster pada Algoritma K- Means dalam Pemetaan E-Government Tahun 2016," *Univ. Res. Colloq.*, pp. 245–250, 2016, [Online]. Available: https://journal.unimma.ac.id/index.php/urecol/article/view/1074/753.

[9] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.

[10] Y. Miftahuddin, S. Umaroh, and F. R. Karim, "Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, Dan Manhattan Dalam Penentuan Posisi Karyawan," *J. Tekno Insentif*, vol. 14, no. 2, pp. 69–77, 2020, doi: 10.36787/jti.v14i2.270.

[11] D. l. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. PAMI-1, no. 2, pp. 224–227, 1979.

[12] T. S. Jaya, "Pengujian Aplikasi dengan Metode Blackbox Testing Boundary Value Analysis," *J. Inform. J. Pengemb. IT*, vol. 03, no. 02, pp. 45–48, 2018.