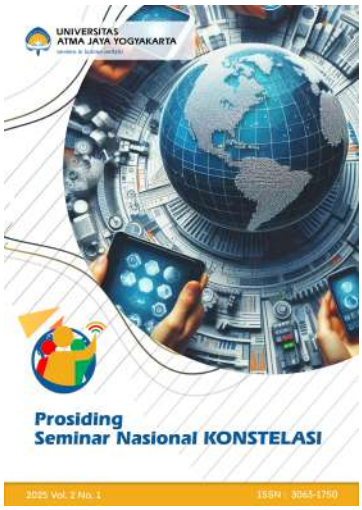
Vol. 2 No. 1 (2025): Juni 2025



Published: 2025-05-30

Articles

Akuisisi Data Pendeteksi Tingkat Stres Manusia berdasarkan Suhu Tubuh, Konduktivitas Kulit, dan Detak Jantung Berbasis IoT (Internet of Things)

Bernadeta Wuri Harini, Rian Bagas Rinaldi

Analisis Upaya Kejahatan Siber Terhadap Pencurian Data Digital Pengguna Internet dan Perangkat Seluler

Ilham Dio Bhakti

Comparison Of Bisecting K-Means And K-Medoids Algorithm In Grouping Junior High School Students Based On Bebras Challenge 2023 Results

Elvina Lorenza Phang

Implementasi Metode SAW pada Sistem Pendukung Keputusan Penentuan Penerima Bantuan Beras BULOG (Studi Kasus: Kantor Desa Alasbuluh)

Ulfa Maulida, Ahmad Lutfi, Ahmad Hamdani

211-219

pdf

Inovasi Digital: Aplikasi GIS Berbasis Web untuk Efisiensi Pelayanan Kesehatan KB di Pulau Timor-Provinsi NTT

Sergius Lamablawa, Patrisius Batarius, Ign. Pricher A.N. Samane, Adri Gabriel Sooai

220-230

pdf

Location Planning of Drone Charging Stations using Geographic Information System (GIS) to Maximize Service Level for on-Demand Food Delivery

Rakan Raihan Ali Mohamad, Bertha Maya Sopha

231-246

pdf

Pemodelan Topik Umpan Balik Orientasi Universitas Menggunakan Deep Learning

Syarif Hidayatullah, Nuhaa Salsabila Shidqiyyah, Dinda Tria Kusuma, Muhammad Fair Nudhin Shidiq, Dian Ayu Fauziah

247-254

pdf

Penerapan PDM dan RCA dalam Perbaikan Proyek Pengadaan Suku Cadang GGW PT. Y di PT. X

Dewi Wulandari, Petrus Setya Murdapa

255-267

pdf

Pengembangan Portal Web Sekolah Menengah Pertama Di Kota Kupang

Fransiskus Felisitus Dalli, Adri Gabriel Sooai, Paskalis Andrianus Nani

268-278

pdf

Rancang Bangun Sistem Administrasi dan Pemesanan Lapangan Futsal Berbasis Web di Indah Futsal

Lucio Marcal Da Silva, Emanuel Jando, Paskalis Andrianus Nani

279-288

pdf

Rancang Bangun Sistem Informasi Rawat Jalan Berbasis Website pada Praktik Mandiri Medika Asembagus

Dila Puspita Dewi Hosip, Ahmad Lutfi, Irma Yunita

289-300

pdf

Sistem Informasi Pengelolaan Rawat Inap pada Klinik BPM (Bidan Praktik Mandiri) Berbasis Web

Carissa Komala Sari Carissa, Ahmad Lutfi, A Hamdani

301-310

pdf

Analisis Kekuatan Desain-Rangka Mesin Palletizer Carton Box dengan Metode Analitis dan Elemen Hingga

Deslana Avinda Krisna Putra, Adhi Setya , Bondan Wiratmoko, Andre Sugijopranoto, Novi Misgi

311-323

pdf

Pengaruh Cahaya Merah Terhadap Produksi Hidrogen Pada Proses Elektrolisis

Mardi Santoso, Purnami, Nurkholis Hamidi

324-332

pdf

Pengembangan Aplikasi KelolaKu Dengan Penerapan Metode SARIMA untuk Optimasi Peramalan dan Pengelolaan Data

Ramadian Ilham Wiguna, Wilfridus Bambang Triadi Handaya, Theresia Devi Indriasari

333-341

pdf

Pengembangan Mesin Press Sepatu Berbasis Hidrolik untuk Meningkatkan Efisiensi Produksi UMKM

Fidelis Gigih Triatmaja, Hieronimus Vikko Purohita, A.Y. Novi Misgi Prabowo Adi, Anton Soesilo

342-355

pdf

Peran Artificial Intelligence ChatGPT dalam Akuntansi: Analisis Bibliometrik

Amalia Khoirunnisa, Sri Hartoko

355-364

pdf

Perancangan Buku Profil Digital Sebagai Upaya Meningkatan Kesadaran Merek Pada Kedai Kopikhan

Swesti Anjampiana Bentri, Briantito Adiwena, Rico Agni Juliano

35-375

pdf

PERANCANGAN DAN PEMBUATAN SISTEM MANAJEMEN DATA STREAMING MUSIC PADA PT. SILOGAN MASINDOTAMA

I Gede Wiarta Sena Sena, Edwin Meinardi

377-390

pdf

Perancangan Sistem Monitoring dan Pengaturan Bahan Pembuatan Kopi pada Mesin Kopi Otomatis

Michael Manchester Mozart, Damar Widjaja

390-400

pdf

Sistem Pendukung Keputusan Sanksi Pelanggaran Siswa Menggunakan Metode SMART Berbasisw Web

Naila Adiba Husein, Ahmad Lutfi, Irma Yunita

401-413

pdf

Strategi Pemeliharaan Preskriptif: Optimalisasi Keandalan Mesin Berbasis Machine Learning Guna Mencegah Terjadinya Downtime Pada Mesin Industri

Inovasi Pencatatan Data Umat GBI Miracle Service Sudirman Berbasis Online Menggunakan QR Code

Magdalena Daisomi Gratia Luna Sari, Patricia Audrey Roellina, Ni Komang Desy Wulandari, Adrial Ignatius Andi Lolo, Generosa Lukhayu Pritalia

477-488

pdf

Kelas Interaktif: Eksplorasi Kreativitas Digital dengan Scratch bagi Siswa/I SMP Pangudi Luhur Sedayu

Ariel Hebron Simanjuntak, Arzetty Cellini Parhusip, Maria Silviany Cahya Serang, Valeria Fevayosa Ginting, Generosa Lukhayu Pritalia

489-499

pdf

Optimalisasi Media Sosial untuk Meningkatkan Engagment pada UMKM Cafe KopiOn Melalui Konten Digital

Gabriel Juansava, Richardo Mario Martin, Putu Gede Garuda Mahendra Oka, Jonathan Damar Subagyo

500-507

pdf

Pelatihan Desain Visual Melalui Canva Pada Pendidikan dan Pelatihan Happy Youth Camp 2025

Nicholas Prakoswa Chandra, Ivan Ong, Jessen Wijaya Agussalim, Flourensia Sapty Rahayu

508-519

pdf

Pengembangan Website dalam Meningkatkan Akses Informasi dan Kegiatan Jemaat Gereja Kristen Jawa Condongcatur

Fanidyasani Atantya, Sion Felix Saragih, Moch Alif Budi Setyawan, Rangga Perwiratama

520-540

pdf

Pengaruh Sosial Media TikTok terhadap Impulsive Buying di Universitas Atma Jaya Yogyakarta Angkatan 2023

Vania Bella Christina, Stefhen Fernandio, Efraim Emmanuel Ransa, Steve Anggana

541-550

pdf

Information

For Readers

For Authors

For Librarians

PROSIDING SEMINAR NASIONAL KONSTELASI

Program Studi Sistem Informasi
Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta
Kampus 3, Gedung Bonaventura
Jln. Babarsari No. 43, Caturtunggal, Kec. Depok, Kabupaten Sleman, DIY 55281
Email: prosidingkonstelasi@uajy.ac.id

Online ISSN: 3063-1750

Comparison of Bisecting K-Means and K-Medoids Algorithms in Clustering Junior High School Students Based on the Results of the 2023 Bebras Challenge

Elvina Lorenza Phang[1], Paulina H. Prima Rosa[2]

[1-2]Information Technology Sanata Dharma University Yogyakarta

E-mail: lorenzaelvina9@gmail.com [1]

Abstract. Improving the quality of education, especially Computational Thinking (CT) skills, is an important priority in the digital era. Bebras Challenge, an international competition to measure Computational Thinking. The results show the diverse abilities of Junior High School students, so that appropriate mentoring methods are needed. This study compares the Bisecting K-Means and K-Medoids algorithms to group students based on the results of the 2023 Bebras Challenge, along with their Indonesian, Mathematics, Science scores, and the duration of competition preparation. The study was conducted through data preprocessing, application of clustering algorithms, and model evaluation using Silhouette Score at the number of clusters k = 2 to k = 10, and running time. The results show two main groups, namely the first group of students with high understanding, fast processing time, and effective preparation. While the second group of students with low understanding due to less than optimal preparation. Bisecting K-Means showed the best performance with a Silhouette Score of 0.623 and an execution time of 0.005 seconds at k = 2. This study provides insights for educators and policy makers to design more effective data-based learning strategies.

Keywords: Computational Thinking , Bisecting K-Means , K-Medoids , Silhouette Score.

1. Introduction

Improvement quality education is priority important in prepare generation young face challenges of the digital era. Assessment results system education by *the Program for International Student Assessment* (PISA) in 2022, showed decline score almost all countries. In Indonesia, the score literacy decrease as many as 12 points , while score literacy mathematics and scores respective scientific literacy decreases as many as 13 points [1]. In 2025, PISA plans add *Computational Thinking* (CT) to in the list of skills evaluated .

Computational Thinking (CT) is ability analytical which includes breakdown problem , design system , and retrieval decision based on principle computer [2]science. Indonesia has promote CT through Bebras Challenge, competition international assessor ability student in think computational . However , the results competition This show diversity ability students , especially at the secondary school level First (junior high school), which requires method appropriate assistance.

Research that uses outcome data *Bebras Challenge* For know level Elementary school students' CT understanding was carried out by Herlina et al. [3] use *K-Means Clustering* .

Various study compare a number of algorithm *clustering* like *K-Means* , *Bisecting K-Means* , *Fuzzy C-Means* , *K-Medoids* , and *Hierarchical* . Researches the show that *Bisecting K-Means* has performance more fast [4]and *K-Medoids* are more robust against *outliers* [5]. However , the comparison second algorithm in context results Bebras Challenge is still on seldom found.

Study This aiming For know results CT data *clustering* of participants Bebras Challenge 2023 junior high school level uses algorithm *Bisecting K-Means* and *K-Medoids* . In addition , also comparing algorithm *Bisecting K-Means* and *K-Medoids* in grouping junior high school students based on results *Bebras Challenge* 2023. Data includes results *Bebras Challenge* and questionnaires distributed to participants *Bebras Challenge* 2023. Evaluation comparison done use *Silhouette Score* For evaluate cluster quality and time execution . Research results expected can give outlook for teachers and services education in designing learning strategies use increase quality student .

2. Research methods

The data used in this study are the results of the 2023 *Bebras Challenge* obtained from *the Bebras* Indonesia competition and questionnaires distributed by *the Bebras* Bureau of Sanata Dharma University to *Bebras Challenge participants* . Both data obtained are in the form of *Microsoft Excel documents* . The data from *Bebras* Indonesia contains the results of the 2023 *Bebras Challenge* in the scouting category, namely the Junior High School (SMP) student level, totaling 7,793 data. The data has 23 attributes containing the participant's name, gender, school of origin, bureau code, total score, time used, and score per question category. While the questionnaire data totals 1,203 data with 13 attributes, containing school level, school name, student name, frequency of participating in *the Bebras Challenge* , how to prepare, total time spent studying, Indonesian, Science, and Mathematics report card scores, impressions, interests, and suggestions.

Both data are preprocessed *in* Data *Mining* with several stages, namely data *cleaning* , data *integration* , data *transformation* using *Robust Scaler* , and data *selection* based on correlation analysis. Then the *Data Mining process is carried out. Mining* by applying *clustering method Bisecting K-Means* and *K-Medoids* . Then the results are evaluated with *Silhouette Score* and *Running Time* .

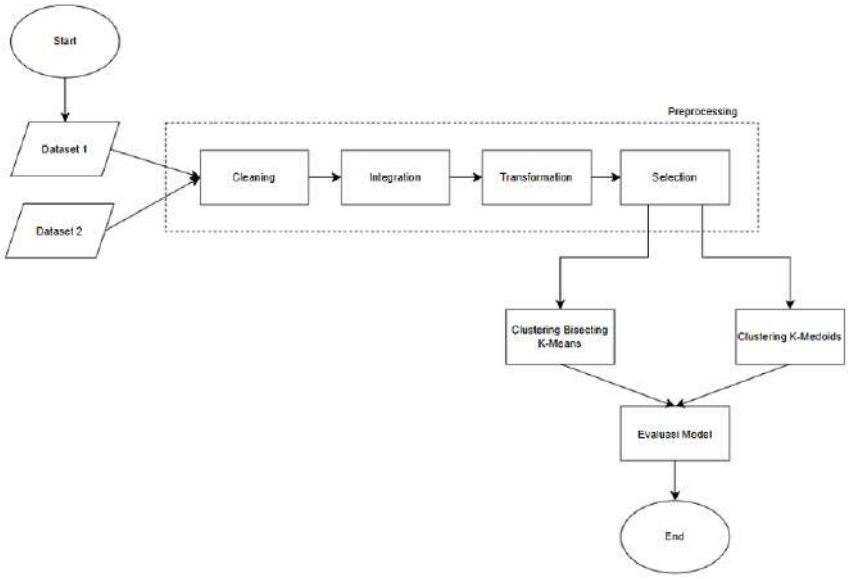


Figure 1 Research Stages

2.1 Data Mining

Data Mining is a technique for obtaining information from data sets and is part of Knowledge Discovery in Database (KDD). KDD includes stages of data analysis to find and identify patterns [6], as shown in Figure 2.

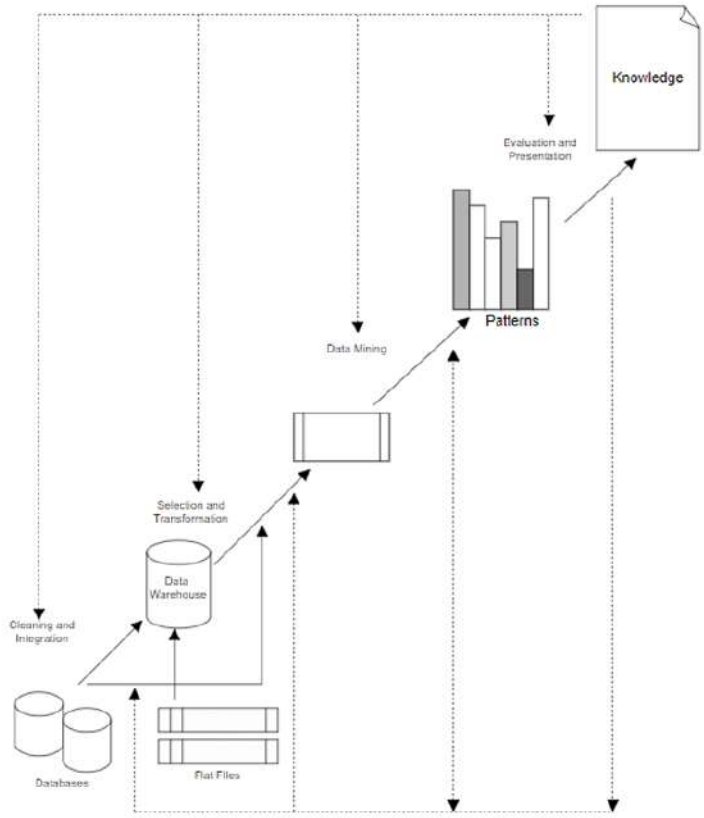


Figure 2. Knowledge Discovery in Databases (KDD)

1. Data Cleaning
   Clean up data with remove noise, duplicate data , or fill in missing value in accordance need .
2. Data Integration
   Merging data from various source with adapt identity such as name or ID.
3. Data Selection
   Selecting relevant data For study with analyze attribute use method like analysis correlation .
4. Data Transformation
   Convert data to the appropriate format For analysis , for example use normalization or scale certain .
5. Data Mining
   Look for pattern or information use method such as K-Means, Bisecting K-Means, K-Medoids, and others .
6. Pattern Evaluation
   Evaluate patterns found use metric such as Silhouette Score, SSE, or Davies-Bouldin Index.
7. Knowledge Presentation
   Serve results and patterns found in form visualization to users , including clustering results.

2.2 Robust Scaler

At the stage KDD transformation , research This use normalization with Robust Scaler for overcome difference range mark between attribute . This technique designed For reduce the influence of outliers with utilizing the median and interquartile range (IQR) in the calculation [7], [8]. The Robust Scaler formula is presented in the equation following .

$$Scaled\_X = \frac{x - Median}{IQR}$$

Information :

- $Scaled\_X$ = new value of *Robust Scaler result*
- X = normalized data using *Robust Scaler*
- median = the median value of a data
- IQR = range between quartile 3 (Q3) and quartile 1 (Q1)

2.3 Analysis Correlation

Correlation analysis is a method for determining the direction, strength, and significance of relationships between variables in a dataset. One technique is the Pearson correlation coefficient, which measures the linear relationship between two variables [9], [10], [11]. The Pearson Correlation formula is presented in the following equation.

$$r = \frac{n * \sum ij - (\sum i) * (\sum j)}{\sqrt{n * \sum i^2 - (\sum i)^2} * \sqrt{n * \sum j^2 - (\sum j)^2}}$$

Information :

- r = *Pearson correlation coefficient* between two variables i and j
- n = amount of data
- $\sum ij$=total number of multiplications between the values in variables i and j
- $\sum i$ = sum of values of variable i
- $\sum j$ = sum of values of variable j

There is guidelines commonly used For assess and understand mark coefficient correlation , which is described in table 1 below This .

Table 1of Correlation Analysis Values

Correlation Value	Interpretation
0.80 – 1.00	Very High Correlation
0.60 – 0.79	High Correlation
0.40 – 0.59	Moderate Correlation
0.20 – 0.39	Correlation Weak
0.00 – 0.19	Very Weak Correlation Even Nothing

2.4 Clustering

Clustering is the process of grouping data based on similarity, where data in one cluster is similar but different from other clusters [5], [12]. The clustering algorithms used in this study are *Bisecting K-Means* and *K-Medoids*.

2.4.1. Bisecting K-Means

Bisecting K-Means is a development of K-Means that combines the Hierarchical principle. This algorithm begins by grouping all data in one cluster, then dividing the cluster into two (k=2) using K-Means until the desired number of clusters is reached [4], [13]. The steps are:

1. Determine number of clusters (k).
2. Using k=2 to divide *the clusters* (initializing 2 *cluster centroids* first).

3. Calculate the similarity of each object data in *the cluster* with both *centroids* using the *Euclidean Distance equation* , then place the object in the nearest *centroid* . The *Euclidean Distance equation* is as follows.

$$D(\mathrm{i},\mathrm{j}) = \sqrt{\sum_{k=1}^{n}\left(x_{ki} - \mathrm{x_{kj}}\right)^2}$$

Information :
- D(i, j) is the distance from data i to the center of *cluster* j.
- x_{yi} is the value of data i in the kth variable
- x_{kj} is the central value of j on the kth variable
- n is the number of observed variables

4. *The cluster* average using the equation below. Then recalculate the similarity of each object data in *the cluster* with the two new *centroids* using the *Euclidean Distance equation* . Steps 2-4 are repeated until the object no longer moves *clusters* .

$$C(t+1) = \frac{1}{Nc_j}\sum_{j_\epsilon c_j} x_j$$

Information:
- $C(t+1)$: new *centroid* (center point) in the next iteration
- Nc_j: the amount of data in a *cluster*

5. Calculating the distance between data and *clusters* to determine which *clusters are still far apart using the Sum of Square Errors* [14]. The SSE formula is as follows.

$$Sum\ of\ Square\ Errors = \sum_{k=1}^{n_j}\left(x_k - \mu_\mathrm{j}\right)^2$$

Information:
- n_j: number of points in *cluster j*
- x_k: kth data point in *cluster j*
- μ_j: *centroid* of *cluster j*

6. Select *the cluster* with the largest amount of data or the largest *SSE* , then divide it into two *clusters* using *K-Means Clustering* .
7. Repeat steps 2-6, until as many *clusters are formed* as the value of k set at the beginning.

2.4.2. K-Medoids

K-Medoids or Partitioning Around Medoids (PAM) is a clustering technique that divides data into several clusters using medoid as the cluster center, not the mean [15]. Medoid is an object that represents the cluster center based on the median [16], [17]. The steps are:

1. Determine the *cluster centers* of k (number *of clusters*)
2. Determine the initial *medoid* randomly k from the total data.
3. With the *Euclidean Distance equation* , the distance of each object and the nearest *cluster* is calculated. Then calculate the total of all distances.
4. Determine new medoids by evaluating candidate *medoids* that provide the minimum distance to each *cluster* .
5. Recalculate the distance of each object in each *cluster* with the new *medoid candidate using the Euclidean Distance equation* , then calculate the total of all distances using the new *medoid* (new total distance).

6. Calculate the total new distance – total old distance to find the total deviation (S). If the result S < 0, the new *medoid* is accepted and updated to produce a new group of k *medoids* . The equation for calculating the total deviation is as follows:

   S = ba

   Information:
   - a : total shortest distance between the object and the initial *medoid*
   - b : total closest distance between object to new *medoid*
7. Repeat steps 4-6 until *the medoids* do not change and produce *clusters* with their respective *cluster members*.

2.5 Silhouette Score

Silhouette Score is method clustering evaluation that measures quality and strength data [18]grouping . This method calculate the average distance between object in one cluster and distance object to another cluster. Values approaching -1 indicate grouping bad , value approaching 1 indicates grouping good , value approaching 0 indicates grouping overlap overlap [6]. The steps for calculating the Silhouette Score are as follows: following :

1. Finding *intra value cluster distance* which is the average distance between the i-th data and other data in the same *cluster* using the equation a(i).

$$a(\mathrm{i}) = \frac{1}{|n| - 1} \sum_{j \in n, j \neq 1} d(i, j)$$

   Information :
   - *intra* -mean value *cluster distance*
   - n is the total number of data in the same *cluster*
   - d(i,j) is the *Euclidean distance* between data i and data j
2. Determining the *inter value cluster distance* which is the average distance of the i-th data with other data in the nearest different *cluster* (minimum) using the equation b(i).

$$b(i) = \min\left(\frac{1}{n(k)} \sum_{j \in n(k)} d(i, j)\right)$$

   Information:
   - *inter* -average value *cluster distance*
   - n(k) is the total number of data in the different *clusters* closest to data i.
   - d(i,j) is the *Euclidean distance* between data i and data j
3. Calculating *silhouette index* with *intra values cluster distance* and *inter cluster distance* is found using the SI(i) equation.

$$SI(i) = \frac{b(i) - a(i)}{max\big(a(i), b(i)\big)}$$

   Information:
   - *intra* -mean value *cluster distance*
   - *inter* -average value *cluster distance*
   - SI(i) is the *Silhouette Score value* of the ith data
4. Get the *Silhouette Score value* of all *clusters* with the S equation.

$$S = \frac{1}{k} \sum_{i=1}^{k} SI(i)$$

   Information:
   - S is the average *Silhouette Score value* of all data points.

- SI_i is the *Silhouette Score value* of the i-th data
- k is the total number of points in the dataset

The range of *cluster strength values based on the Silhouette Score* evaluation results can be seen in table 2.

Table 2 Silhouette Score Evaluation Value Range

Range	Interpretation
0.71 - 1.00	*The clusters* formed produce strong structures.
0.51 – 0.70	*The clusters* formed produce good structures.
0.26 – 0.50	*The clusters* formed produce weak structures.
≤ 0.25	*The clusters* formed are unstructured

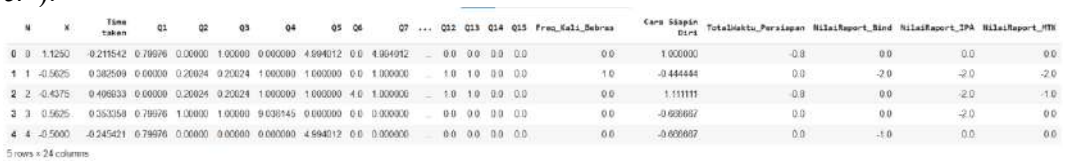
3. Results and Analysis

In this section, the research results are explained and a comprehensive discussion is provided. The research results can be presented in the form of images, graphs, tables and others that make it easier for readers to understand them [2], [5] . The discussion can be divided into several sub-chapters.

This study used data from the 2023 *Bebras Challenge* (raising category) and questionnaire data from *the Bebras Bureau* of Sanata Dharma University.

3.1 Preprocessing Stage

- Data *cleaning* (removing duplicate data, converting values in the " *Time taken* " attribute to seconds, changing hyphens *"-"* to 0 *,* removing characters, dots, numbers in attributes containing student names).
- Data *integration* (combining) both datasets are based on attributes containing student names , resulting in 36 attributes from a dataset).
- Data *transformation* (changing attribute names , initialization) student name attribute , conversion a number of attribute categorical become numeric *,* do normalization *Robust Scaler*).



	N	K	Time taken	Q1	Q2	Q3	Q4	Q5	Q6	Q7	...	Q12	Q13	Q14	Q15	Frec_Kali_Bebras	Cara Siapin Diri	TotalWaktu_Persiapan	NilaiRaport_Bind	NilaiRaport_IPA	NilaiRaport_MTK
0	0	1.1250	-0.211542	0.79976	0.00000	1.00000	0.000000	4.994012	0.0	4.994012	...	0.0	0.0	0.0	0.0	0.0	1.000000	-0.8	0.0	0.0	0.0
1	1	-0.5625	0.382509	0.00000	0.20024	0.20024	1.000000	1.000000	0.0	1.000000	...	1.0	1.0	0.0	0.0	1.0	-0.444444	0.0	-2.0	-2.0	-2.0
2	2	-0.4375	0.406833	0.00000	0.20024	0.20024	1.000000	1.000000	4.0	1.000000	...	1.0	1.0	0.0	0.0	0.0	1.111111	-0.8	0.0	-2.0	-1.0
3	3	0.5625	0.353358	0.79976	1.00000	1.00000	9.036145	0.000000	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	-0.666667	0.0	0.0	-2.0	0.0
4	4	-0.5000	-0.245421	0.79976	0.00000	0.00000	0.000000	4.994012	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	-0.666667	0.0	-1.0	0.0	0.0

5 rows × 24 columns

Figure 3 Example of Robust Scaler Normalization Results

- Data *selection* (doing analysis correlation , data selection from 36 attributes into 24 attributes).

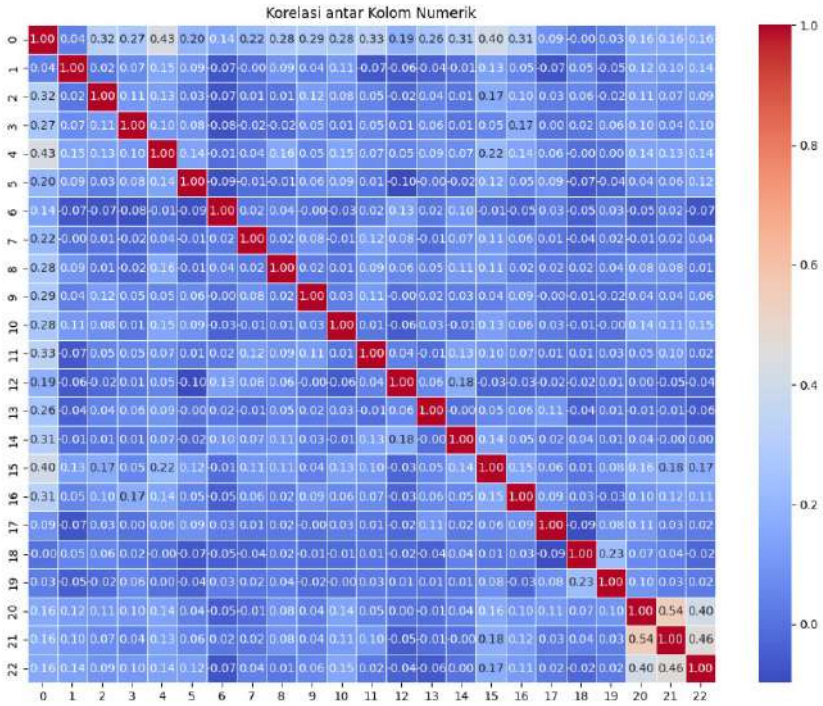


Figure 4 Correlation Analysis

Analysis results correlation show that correlation between attribute dominated with correlation weak until moderate , which means attributes the No own strong relationship . This is can help in the process *of clustering* Because more attributes diverse and not very tied One each other.

3.2 Clustering Testing

Election best k amount seen from mark *Silhouette Score* and running time of each algorithm . Trial done use number k = 2 to 10.

Testing First use *Bisecting K-Means Clustering.* Figure 5 shows mark *Silhouette Score* (y- axis) for every k value (x- axis). While Figure 6 displays mark *running time* (y- axis) for every k value (x- axis).

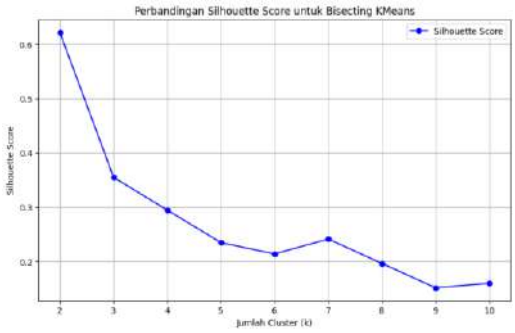


Figure 5 Comparison Results Graph of Silhouette Score Bisecting K-Means

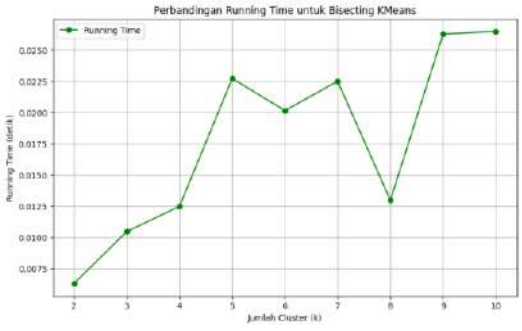


Figure 6 Comparison Results Graph of Running Time Bisecting K-Means

Table 3 Results Silhouette Score testing and Bisecting K-Means running time

k	Silhouette Score	Running Time
2	0.623483	0.005587
3	0.356208	0.005317
4	0.296267	0.006277

k	Silhouette Score	Running Time
5	0.243623	0.009165
6	0.236400	0.013957
7	0.192294	0.016562
8	0.205754	0.011388
9	0.165149	0.012815
10	0.173626	0.012953

Testing second use *K-Medoids Clustering.* Figure 7 shows mark *Silhouette Score* (y- axis) for every k value (x- axis). While Figure 8 displays mark *running time* (y- axis) for every k value (x- axis).

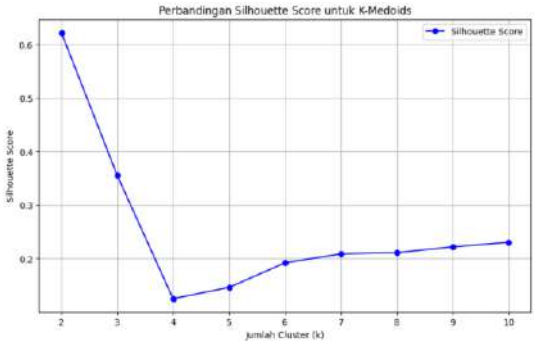


Figure 7 Comparison Results Graph of Silhouette Score K-Medoids

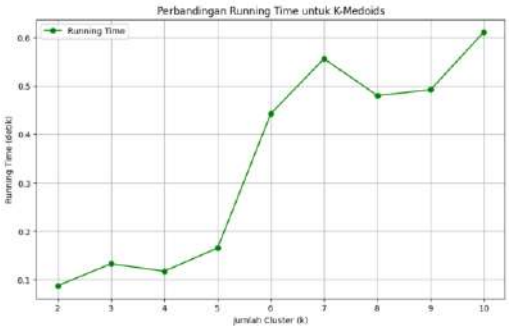


Figure 8 Comparison Results Graph of K-Medoids Running Time

Table 4 Results Silhouette Score and K-Medoids running time testing

k	Silhouette Score	Running Time
2	0.623483	0.040914
3	0.356208	0.059698
4	0.126033	0.071001
5	0.147244	0.133113
6	0.197766	0.246575
7	0.210053	0.247694
8	0.215018	0.334253
9	0.224597	0.383522
10	0.232982	0.472803

Based on images and tables the above test , the results testing shows k=2 as results best with a Silhouette Score of 0.623 and a running time of 0.006 for Bisecting K-Means, and 0.041 for K-Medoids. Both algorithm show that improvement number of clusters (k) decreases quality Silhouette Score evaluation , while running time increases along increase k value .

3.2.1 *Cluster* Results Using *Bisecting K-Means*

Based on the results of the cluster analysis produced by *Bisecting K-Means clustering* , *cluster* 0 consists of 70 participants with higher total scores and processing time (average 30 minutes) compared to cluster 1 which consists of 538 participants (average 33 minutes). Although cluster 1 has a slightly higher average score for some questions, cluster 0 tends to be filled by participants with higher understanding and faster processing time.

Profiling several categorical attributes shows that the majority of participants in both clusters have only participated in the Bebras Challenge once. Cluster 0 is dominated by participants who are more experienced and rely more on teacher guidance, while cluster 1 uses more flexible preparation methods, such as self-study. Participants in cluster 0 also tend to allocate more preparation time and have higher report card scores in the 81-90 and >90 categories, especially in Science and Mathematics subjects. Cluster 1 is more dominant in the 71-80 and 81-90 scores in Science, while cluster 0 excels in the >90 scores. Overall, cluster 0 shows participants with better understanding and preparation.

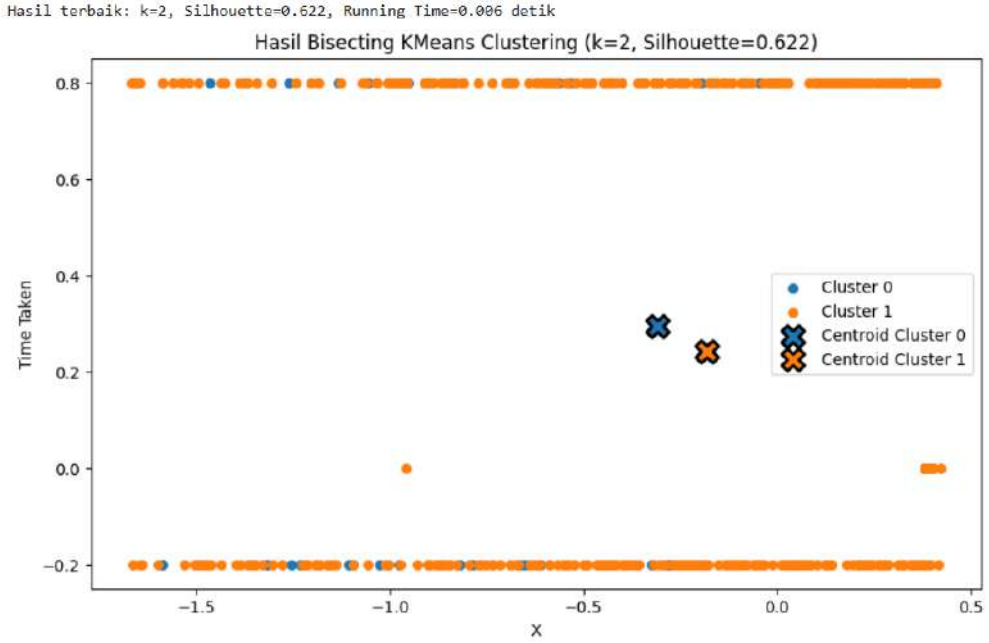


Figure 9 Bisecting K-Means Clustering Results

3.2.2 *Cluster* Results Using *K-Medoids*

Based on the results of the cluster analysis produced by *K-Medoids clustering* , Cluster 0 consists of 538 participants with lower total scores and longer completion times compared to cluster 1 which contains 70 participants. Although cluster 1 has a higher average score and faster completion times, cluster 0 includes participants with varying experiences, including those who have participated in the Bebras Challenge more than 3 times. Cluster 1 is dominated by participants who study more often with teachers, while cluster 0 is more flexible with various preparation methods.

Profiling of several categorical attributes shows that Cluster 0 has more participants who prepare for 1-3 hours, while cluster 1 has more who prepare for more than 5 hours. Based on report card scores, cluster 1 excels with a higher percentage in the 81-90 and >90 categories, while cluster 0 is more in the 81-90 range. For Science, cluster 1 excels in the high score category (>90), while cluster 0 dominates in the 81-90 range. In Mathematics, cluster 1 participants have more scores >90 and slightly more who have scores below 70. Overall, cluster 1 shows participants with higher scores and longer preparation.

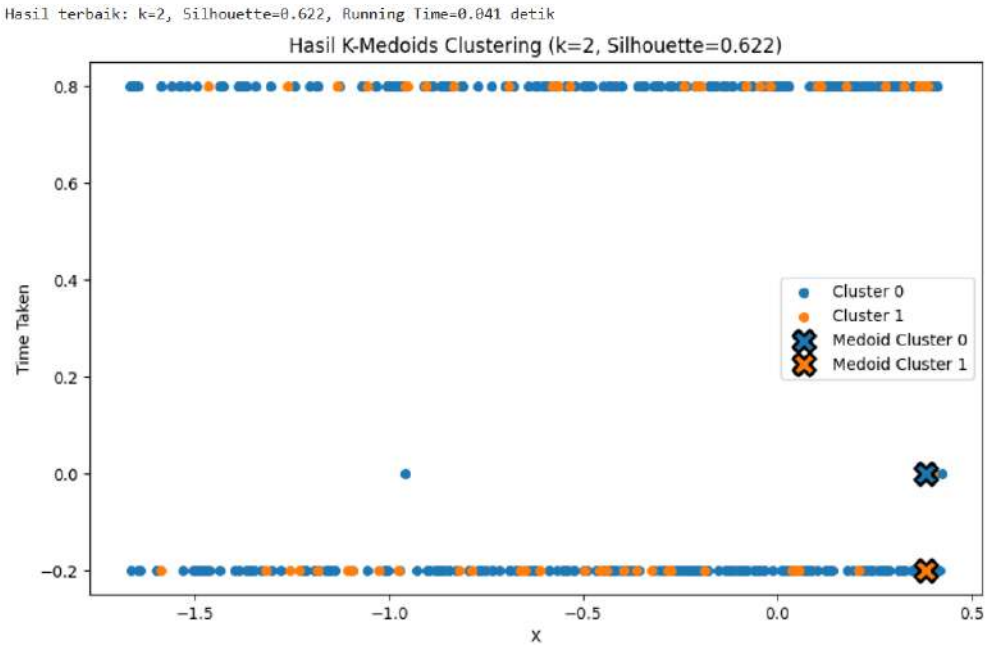


Figure 10K-Medoids Clustering Results

4. Conclusion

Based on research that has been done , can concluded that the Computational Thinking Bebras Challenge 2023 data clustering uses Bisecting K-Means and K-Medoids algorithms produce two clusters with difference characteristics understanding participants . Group First show understanding more tall with time workmanship more fast , value more report cards high , and more preparation effective with the teacher, while group second own better understanding low consequence lack of preparation effective and time learn more little . Comparison algorithm based on *Silhouette Score* and *Running Time* show that Bisecting K-Means algorithm is proven more Good in do clustering because own time more execution fast , namely 0.006 seconds compared to K-Medoids which require time 0.041 seconds . Both algorithm This produce mark the same Silhouette Score evaluation which is 0.623 at k=2.

As a suggestion for study Next , it is recommended For try other clustering algorithms such as DBSCAN or Agglomerative Clustering for more results optimal. Use method evaluation addition such as the Davies-Bouldin Index and the Dunn Index can give outlook more about Cluster quality . Calculation method distance like Manhattan can also under consideration For compared to with Euclidean. In addition , this dataset Can used For classification use identify factors that influence success participant in Bebras Challenge.

5. Reference

[1] *PISA 2022 Results (Volume I)* . in PISA. OECD, 2023. doi: 10.1787/53f23881-en.
[2] J.M. Wing, *Computational Thinking* . IEEE, 2011.
[3] Herlina and Z. Ernaningsih, “Implementation of K-Means Clustering for Analysis of Elementary School Students’ Computational Thinking Understanding Level,” *BUDIDARMA INFORMATICS MEDIA JOURNAL* , vol. 7, no. 3, pp. 1405–1413, 2023, doi: 10.30865/mib.v7i3.6132.
[4] E. Seniwati, A. Sidauruk, Haryoko, and A. Lukman, "Clustering Performance Between K-means and Bisecting K-means for Students Interest in Senior High School," *Building of Informatics, Technology and Science (BITS)* , vol. 5, no. 1, June. 2023, doi: 10.47065/bits.v5i1.3624.

[5] E. Luthfi, A. Wahyu Wijayanto, and P. Statistika, "Comparative analysis of hierarchical, k-means, and k-medoids clustering methods in grouping the Indonesian human development index," *Jurnal FEB* , no. 4, pp. 761–773, 2021, [Online]. Available: http://journal.feb.unmul.ac.id/index.php/INOVASI

[6] J. Han, M. Kamber, and J. Pei, "Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)," 2011.

[7] A. Ramsauer, P. M. Baumann, and C. Lex, "The Influence of Data Preparation on Outlier Detection in Driveability Data," *SN Comput Sci* , vol. 2, no. 3, May 2021, doi: 10.1007/s42979-021-00607-7.

[8] D. Sarkar, R. Bali, and T. Sharma, *Practical Machine Learning with Python* . Apress, 2018. doi: 10.1007/978-1-4842-3207-1.

[9] T. Budiwati, A. Budiyono, W. Setyawati, and A. Indrawati, "PEARSON CORRELATION ANALYSIS FOR CHEMICAL ELEMENTS OF RAINWATER IN BANDUNG," *Jurnal Sains Dirgantara* , vol. 7, no. 2, 2010.

[10] F. Mayang Sari, R. Nur Hadiati, and W. Perinduri Sihotang, "Pearson Correlation Analysis of Total Population and Number of Motorized Vehicles in Jambi Province," *Journal of Statistics, University of Jambi* , vol. 2, no. 1, 2023, doi: 10.22437/multiproximity.v2i1.25568.

[11] WR Safitri, "PEARSON CORRELATION ANALYSIS IN DETERMINING THE RELATIONSHIP BETWEEN DENGUE FEVER INCIDENTS AND POPULATION DENSITY IN SURABAYA CITY IN 2012 - 2014," *Journal of STIKES Pemkab Jombang* , 2014.

[12] I. Kamila, U. Khairunnisa, and Mustakim, "Comparison of K-Means and K-Medoids Algorithms for Clustering," *Scientific Journal of Information Systems Engineering and Management* , vol. 5, no. 1, pp. 119–125, 2019.

[13] S. Dwididanti and DA Anggoro, "Comparative Analysis of Bisecting K-Means and Fuzzy C-Means Algorithms on Credit Card User Data," *Emitter: Jurnal Teknik Elektro* , vol. 22, no. 2, pp. 110–117, Aug. 2022, doi: 10.23917/emitor.v22i2.15677.

[14] R. Nainggolan, R. Wargain-Angin, E. Simarmata, and AF Tarigan, "Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method," in *Journal of Physics: Conference Series* , Institute of Physics Publishing, Dec. 2019. doi: 10.1088/1742-6596/1361/1/012015.

[15] F. Faisal, L. Aliska Giopani, F. Ma'idatul, Z. Cindya Dwynne, S. Syahidatul Helma, and Mustakim, "Comparison of K-Means and K-Medoids Algorithms for Temperature Clustering in Riau Province," *Indonesian Journal of Informatic Research and Software Engineering* , vol. 2, no. 2, pp. 128–134, 2022.

[16] S. Bahri, D. Marisa Midyanti, and P. Correspondence, "APPLICATION OF K-MEDOIDS METHOD FOR DROPOUT POTENTIAL STUDENT GROUPING," *Journal of Information Technology and Computer Science (JTIIK)* , vol. 10, no. 1, pp. 165–172, 2023, doi: 10.25126/jtiik.2023106643.

[17] MD Doi, A. Rusgiyono, and T. Wuryandari, "k-MEDOIDS ANALYSIS WITH INDEX VALIDATION ON 3T REGIONAL HDI IN INDONESIA," *Gaussian Journal* , vol. 12, no. 2, pp. 178–188, Jul. 2023, doi: 10.14710/j.gauss.12.2.178-188.

[18] Y. Januzaj, E. Beqiri, and A. Luma, "Determining the Optimal Number of Clusters using Silhouette Score as a Data Mining Technique," *International journal of online and biomedical engineering* , vol. 19, no. 4, pp. 174–182, 2023, doi: 10.3991/ijoe.v19i04.37059.