Jurnal Mantik, 6 (2) (2022) 1542-1550
Published by:Institute of Computer Science (IOCS)
Jurnal Mantik
Journal homepage: www.iocscience.org/ejournal/index.php/mantik/index

Clustering Pop Songs Based On Spotify Data Using K-Means And K-Medoids Algorithm

[1]Novia Ayu Privandhani, [2]Sulastri

[12]Information System, Information Technology and Industry Faculty, Stikubank University, Jl. Tri Lomba Juang, Semarang, 50241, Indonesia

E-mail: [1]noviaayuprivandhani@mhs.unisbank.ac.id, [2]sulastri@edu.unisbank.ac.id

ARTICLE INFO

Article history:
Received: Jun 17, 2022
Revised: Jun 29 , 2022
Accepted: Jul 04, 2022

Keywords:
Data Mining,
Clustering,
K-Means,
K-Medoids,
Spotify

ABSTRACT

The development of music service technology is currently making it easier to listen to songs. One of them is the Spotify application. Services on music attributes are random, such as Danceability, Energy, Acousticness, Instrumentalness, Liveness, Loudness, Speechiness, Valence, and Tempo. The purpose of this study is to compare the results of clustering on the K-Means and K-Medoids algorithms using the Rstudio tools. The results of this comparison obtain the optimal number of clusters and obtain high, medium, and low cluster results. The results of the K-Means calculation are based on the average in cluster 1 the highest is Tempo with a value of 118 in cluster 2 the highest is Tempo with a value of 125, and in cluster 3 the highest is Tempo with a value of 123. The calculation of K-Medoids is based on the average in cluster 1 the highest is Tempo with a value of 129, in cluster 2 the highest is Tempo with a value of 122, and in cluster 3 the highest is Tempo with a value of 110.

1. Introduction

Rapid technological developments provide music streaming services that are increasingly complete and easily accessible from various platforms. One of them is the Spotify application. Spotify provides a variety of music, podcast, and video services that provide millions of access to songs and other content. Spotify is a big platform and has a lot of users, of course, it needs analysis to get better and increase competition on other platforms. This data uses Spotify's public data based on the genre of Indonesian Pop songs based on 9 audiotrack feature attributes, such as Danceability, Energy, Acousticness, Instrumentalness, Liveness, Loudness, Speechiness, Valence, and Tempo. Each song has a different level of audio features. The music attributes provided by Spotify are random, it is necessary to cluster each song attribute and optimize it because people tend to want to listen to the songs they want to hear[1]. Based on these problems, it is necessary to analyze the attributes of the audio track features using the K-Means and K-Medoids algorithm methods. The purpose of this method is to cluster the database on the audiotrack feature attributes to produce several clusters.

Based on the previous analysis, the K-Means algorithm was used for the process of grouping active students with the elbow method. The results are used as a basis for determining the same cluster in different amounts of data [2]. The previous analysis of the K-Medoids algorithm was used to group new students to determine the pattern of study program selection. These results are based on the testing process using the Silhouette Coefficient method, with the best value of 0,690754 with a total of 3 clusters and a total of 15 data [3]. The K-Means and K-Medoids algorithm methods are used to compare the results with the Spotify genre. By using the Global Top 50 data and calculating 3 clusters, cluster 1 contains 2,833 members, cluster 2 contains 21 members [4]. The grouping of potential forest and land fires based on hotspot has a Silhouette Coefficient global K-Means value of 0,558. While the K-Medoids are 0,529 [5]. Data collection of insurance products of national companies, using 3 attributes, namely premium, number of customers, and year of release for each product. The test results using the K-Means method with the smallest value is K = 5, namely 0,118. While the K-Medoids algorithm, the smallest value is K = 2, which is 0,027 [6]

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

The data collection used is based on Spotify data for the Indonesian Pop song genre with a total of 3,259 records. Data collection techniques are carried out through the scraping process. Web Scraping is a data collection method that has been provided by certain platforms via the internet [7]. A web scraper is generally a page on the web in a markup language such as HTML or XHTML. The document is used to analyze using scraping tools and collect data via the web [8]. The comparison of K-Means and K-Medoids resulted in groups based on the highest, medium, and low clusters.

2. Method

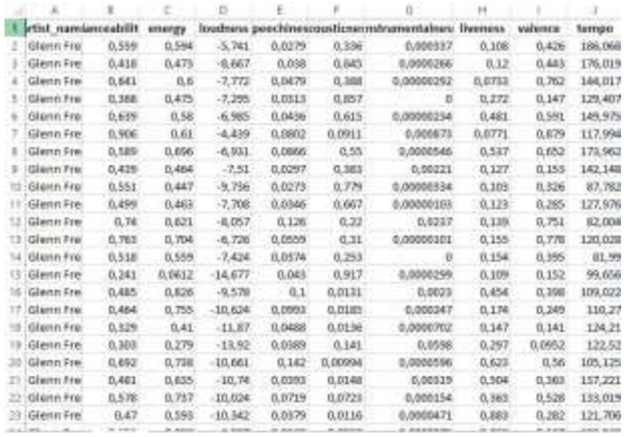
Data collection results from grouping based on attributes of Indonesian pop genre song data using Spotify Developer data obtained through Web Scraping on Rstudio, after that the data will be processed through Data Mining Clustering using K-Means and K-Medoids algorithms to determine the highest cluster results, medium, and low. Data Mining is the process of finding patterns and knowledge from large amounts of data that helps make decisions to use. The pattern is analyzed to find new knowledge that is used to solve a strategic problem in the form of future predictions [9]. Data mining has several signs, such as knowledge is still hidden in data chunks [10].

Clustering is a data mining technique used to find cluster characteristics of large data. Clustering groups large amounts of data into clusters which later in each cluster contains the same or similar data. One example is database technology, data warehouse, statistics, machine learning, computing, and others [11]. The K-Means algorithm is an iteration clustering algorithm on the data partition into the number of K clusters that have been set at the beginning [12]. The K-Means algorithm is implemented because it is relatively fast, adaptable, and generally used in practice [12]. K-Medoids algorithm or Partitioning Around Medoids (PAM) is an algorithm that performs cluster grouping by using a representative object (medoid) as the center of the cluster for each cluster [13]. The K-Medoids algorithm is usually used to overcome the weakness that exists in the K-Means algorithm, because the K-Means is very sensitive to outliers and the object is very far from the majority of other data, so if placed in a cluster, this kind of data can deviate from the average value (mean) of the cluster [14].

This comparative analysis aims to compare the results on the attribute data of audiotrack features based on the genre of Indonesian Pop songs by grouping the K-Means and K-Medoids algorithms. There are several stages in the Clustering process, namely using the data analysis method according to CRISP-DM (Cross Industry Standard Process for Data Mining) [15] which is divided into 6 stages, namely :

a. Business Understanding Stage
   Problem understanding based on audio feature attributes. This stage is needed for audiotrack feature analysis to find out the results of each cluster to compare the results of K-Means and K-Medoids.
b. Data Understanding Stage
   This stage is carried out for initial data collection, then will analyze the data and understand all the contents of the data based on the Data API on Spotify Developer. The technique used to get Spotify data is web scraping. The web scraper technique takes data that has been published on the web [16]. Then it will be saved in a soft file in XLSX format. The data collected is in the form of attribute data for the audiotrack feature with the Indonesian pop song genre, The data collected is 3.259 data and stored in Microsoft Excel. Figure 1 shows a snippet of the initial dataset before it is processed.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.



	A	B	C	D	E	F	G	H	I	J
1	rtist_nam	anceabilit	energy	loudness	peechines	cousticnes	nstrumentalnes	liveness	valence	tempo
2	Glenn Fre	0,559	0,594	-5,741	0,0279	0,336	0,000337	0,108	0,426	186,068
3	Glenn Fre	0,418	0,473	-8,667	0,038	0,645	0,0000266	0,12	0,443	176,019
4	Glenn Fre	0,641	0,6	-7,772	0,0479	0,388	0,00000292	0,0733	0,762	144,017
5	Glenn Fre	0,388	0,475	-7,295	0,0313	0,857	0	0,272	0,147	129,407
6	Glenn Fre	0,639	0,58	-6,985	0,0436	0,615	0,00000234	0,481	0,591	149,975
7	Glenn Fre	0,906	0,61	-4,439	0,0802	0,0911	0,000873	0,0771	0,879	117,994
8	Glenn Fre	0,589	0,696	-6,931	0,0866	0,55	0,0000546	0,537	0,652	173,962
9	Glenn Fre	0,439	0,484	-7,51	0,0297	0,383	0,00221	0,127	0,155	142,148
10	Glenn Fre	0,551	0,447	-9,756	0,0273	0,779	0,00000934	0,103	0,326	87,782
11	Glenn Fre	0,499	0,463	-7,708	0,0346	0,667	0,00000103	0,123	0,285	127,976
12	Glenn Fre	0,74	0,621	-8,057	0,126	0,22	0,0237	0,139	0,751	82,008
13	Glenn Fre	0,763	0,704	-6,726	0,0559	0,31	0,00000101	0,155	0,778	120,028
14	Glenn Fre	0,518	0,559	-7,424	0,0374	0,253	0	0,154	0,395	81,99
15	Glenn Fre	0,241	0,0612	-14,677	0,043	0,917	0,0000299	0,109	0,152	99,656
16	Glenn Fre	0,485	0,826	-9,578	0,1	0,0131	0,0023	0,454	0,398	109,022
17	Glenn Fre	0,484	0,755	-10,624	0,0993	0,0185	0,000347	0,174	0,249	110,27
18	Glenn Fre	0,329	0,41	-11,87	0,0488	0,0136	0,0000702	0,147	0,141	124,21
19	Glenn Fre	0,303	0,279	-13,92	0,0389	0,141	0,0598	0,297	0,0952	122,52
20	Glenn Fre	0,692	0,738	-10,661	0,142	0,00994	0,0000596	0,623	0,56	105,125
21	Glenn Fre	0,481	0,835	-10,74	0,0393	0,0148	0,00319	0,504	0,363	157,221
22	Glenn Fre	0,578	0,737	-10,024	0,0719	0,0723	0,000154	0,363	0,528	133,019
23	Glenn Fre	0,47	0,593	-10,342	0,0379	0,0116	0,0000471	0,883	0,282	121,706

Figure 1. Initial dataset

Information on each attribute column of the Indonesian pop song dataset can be seen in table 1.

TABLE 1
Atribute Column

No	Attribute	Description
1	Danceability	Refers to how ideal a song is to make listeners dance
2	Energy	Measuring the energy of a number
3	Loudness	Perception of song or sound volume
4	Speechiness	Refers to the density of words in the song
5	Acousticness	Measures the extent of effect and distortion on each instrument
6	Instrumentalness	Refers to the variety of instruments used in one song
7	Liveness	Measuring music production with roles when songs are performed live
8	Valence	Refers to the emotions that appear in the song
9	Tempo	Scales that use the same notes as the standard major but in a new and different way

c. Data Preparation Stage
This stage is done to prepare the data. The data obtained is still in the form of raw data, therefore this stage is carried out to consider the completeness of the data and the selection of fields.

d. Modeling Stage
This stage is carried out to determine the appropriate modeling technique. This study uses the K-Means and K-Medoids algorithms with several techniques to solve these problems. Figure 2 shows the modeling stage.

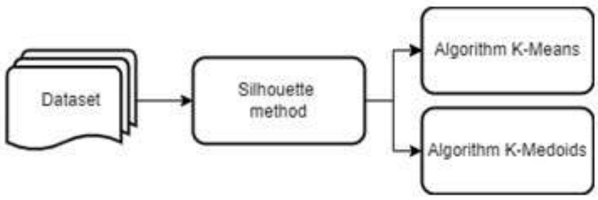


Figure 2. Modeling Stage

Figure 1 explains that the dataset is then clustered optimally using the Silhouette method. By using 3 clusters to determine the results of high, medium, and low clusters on the K-Means and K-Medoids algorithms.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

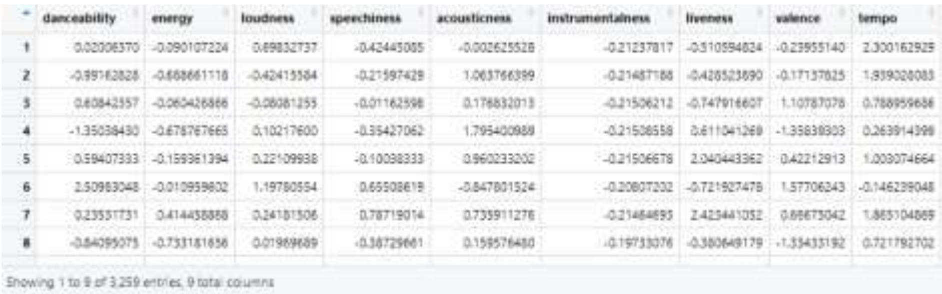
e. Evaluation Stage
   At this stage, an evaluation of the modeling technique used is carried out to determine whether the modeling technique chosen is appropriate and to see if there are problems in the research that are not resolved properly. Then conclude from the results that have been completed. In this study, no evaluation stage was carried out.
f. Deployment Stage
   At this stage, the preparation of the report presents the results of the modeling technique that has been used and then explains it to the parties concerned. In this study, it was not used at the deployment stage.

3. Results and Discussion

The implementation of the K-Means and K-Medoids algorithms using the Rstudio tools has several stages, the first stage is to install and load the package, then data preparation, look for the optimal K cluster, and the last is the execution of K-Means and K-Medoids.

3.1 Data Preparation

Preparing the dataset used in the calculation process Rstudio has several stages such as importing the data, cleaning the data, and selecting the variables used in the clustering process. Figure 3 shows data that can be used to perform cluster calculations with the file name "audio fix".



	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence	tempo
1	0.02006370	-0.090107224	0.69832737	-0.42445085	-0.002625528	-0.21237817	-0.310594824	-0.23955140	2.300162929
2	-0.99162828	-0.668661118	-0.42415584	-0.21597429	1.063766399	-0.21487188	-0.428523690	-0.17137625	1.939028083
3	0.60842557	-0.060426866	-0.08081255	-0.01162598	0.176832013	-0.21506212	-0.747916607	1.10787078	0.788959686
4	-1.35038430	-0.678767665	0.10217600	-0.35427062	1.795400989	-0.21508558	0.611041269	-1.35838303	0.263914399
5	0.59407333	-0.159361394	0.22109938	-0.10038333	0.960233202	-0.21506678	2.040443362	0.42212913	1.003074664
6	2.50983048	-0.010959802	1.19780554	0.65508619	-0.847801524	-0.20807202	-0.721927476	1.57706243	-0.146239048
7	0.23551731	0.414458888	0.24181506	0.78719014	0.735911276	-0.21464693	2.423441052	0.66673042	1.865104869
8	-0.84095075	-0.733181656	0.01969689	-0.38729661	0.159576480	-0.19733076	-0.380649179	-1.33433192	0.721792702

Showing 1 to 9 of 3,259 entries, 9 total columns

Figure 3. A file named "audio fix"

3.2 Optimal K Cluster

The next stage is to find the optimal K using data that has been processed through data preparation beforehand. Finding the optimal K on the K-Means and K-Medoids algorithm using the Silhouette Coefficient method. Figures 4 and 5 show the K optimal results for K-Means and K-Medoids.

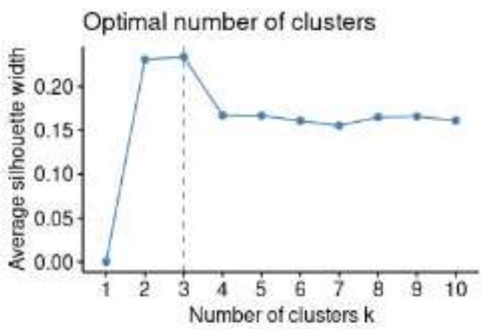


Figure 4. Visualization of optimal K results on K-Means

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

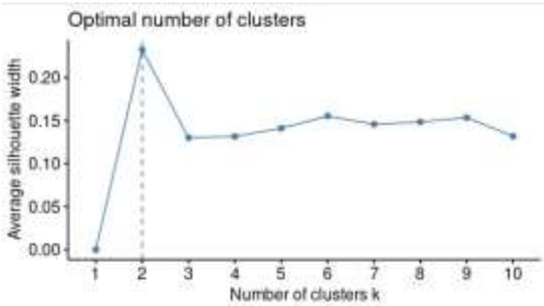


Figure.5 Visualization of optimal K results on K-Medoids

In optimal K results, the K-Means algorithm uses 3 clusters, while the K-Medoids algorithm uses 2 clusters. However, because the purpose of this study was to obtain high, medium, and low cluster results, the calculation process was carried out using the 3 clusters.

3.3 K-Means Execution

The next step is to calculate the K-Means algorithm because the optimal K has been determined, namely K=3, Figure 6 shows the results of cluster membership and cluster means at Rstudio.

```
> print(final)
K-means clustering with 3 clusters of sizes 277, 1208, 1774

Cluster means:
  danceability     energy   loudness speechiness acousticness instrumentalness
1  0.009106745  0.4538720 -0.2045984  1.66494971   -0.2703396      -0.14565868
2 -0.595334145 -0.9610734 -0.5579030 -0.30979284    0.8183202       0.01839052
3  0.403969041  0.5835705  0.4118493 -0.04901991   -0.5150207       0.01022080
    liveness    valence       tempo
1  2.1762505  0.2340674  0.00727913
2 -0.2142857 -0.8851929 -0.15161172
3 -0.1938919  0.5662211  0.10210295
```

Figure 6. Cluster membership and cluster means

Figure 6 describes the grouping of 3 clusters, cluster 1 each with 277 members, cluster 2 each with 1.208 members, and cluster 3 each with 1.774 members. Figure 7 shows the results of the visualization of the K-Means algorithm. Cluster 1 is marked in red, cluster 2 is marked in green, and cluster 3 is marked in blue.

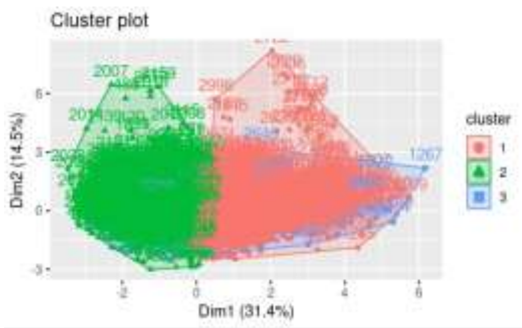


Figure7. visualization of K-Means algorithm

The final result of clustering is the process of calculating the average of each cluster. Figure 8 shows the final result of the K-Means algorithm.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

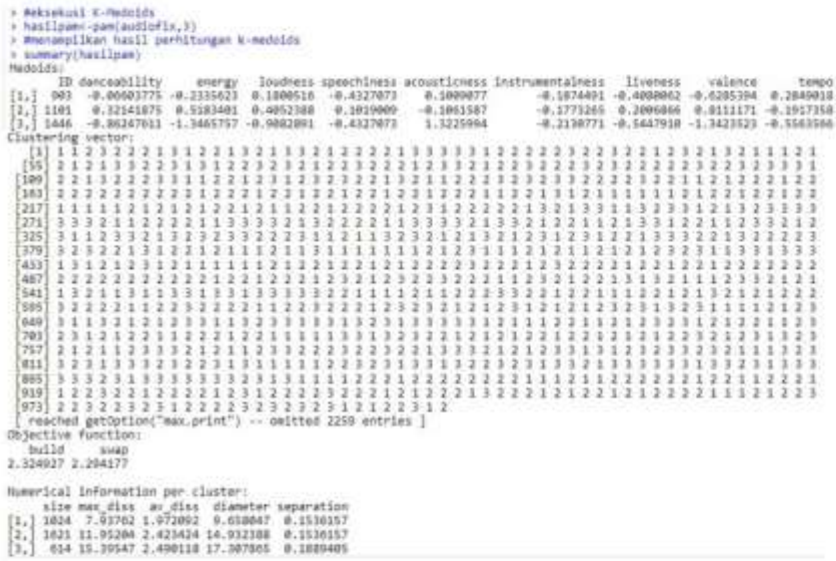
```
> audio[2:10]%>%
+   mutate(cluster = final$cluster)%>%
+   group_by(cluster)%>%
+   summarise_all("mean")
# A tibble: 3 x 10
  cluster danceability energy loudness speechiness acousticness instrumentalness liveness valence tempo
    <int>        <dbl>  <dbl>    <dbl>       <dbl>        <dbl>            <dbl>    <dbl>   <dbl> <dbl>
1       1        0.481  0.422    -9.06      0.0381        0.560          0.00733    0.162   0.274  118.
2       2        0.607  0.736    -6.52      0.0556        0.188          0.00620    0.197   0.624  125.
3       3        0.516  0.614    -8.90      0.0398        0.407          0.697      0.161   0.461  123.
>
```

Figure8. Final K-Means

Figure 8 explains that in cluster 1 it is Tempo because it has the highest value among other audio features reaching a value of 118, in cluster 2 it is Tempo because it has the highest value among other audio features reaching a value of 125, in cluster 3 it is Tempo because it has the highest value. High among other audio features reaches a value of 123.

3.4 K-Medoids Execution

The next step is the calculation of the K-Medoids algorithm because K optimal has been determined K=3, then the calculation of the K-Medoids algorithm can be carried out, Figure 9 displays the medoid representation, clustering vector, and numerical information for each cluster.



```
> #eksekusi K-Medoids
> hasilpam<-pam(audiofix,3)
> #menampilkan hasil perhitungan k-medoids
> summary(hasilpam)
Medoids:
       ID danceability     energy   loudness speechiness acousticness instrumentalness   liveness    valence      tempo
[1,]  903  -0.00603775 -0.2335623  0.1800516  -0.4327073    0.1009077       -0.1874491 -0.4008062 -0.6285394  0.2849018
[2,] 1101   0.32141875  0.5183401  0.4052388   0.1019009   -0.1061587       -0.1773265  0.2006066  0.8111171 -0.1917358
[3,] 1446  -0.86247013 -1.3465757 -0.9082091  -0.4327073    1.3225994       -0.2130771 -0.5447910 -1.3423523 -0.5563566
Clustering vector:
[ reached getOption("max.print") -- omitted 2259 entries ]
Objective function:
   build     swap
2.324927 2.204177

Numerical information per cluster:
     size max_diss  av_diss  diameter separation
[1,] 1024  7.93762 1.972092  9.618047  0.1536157
[2,] 1621 11.95204 2.423424 14.932188  0.1536157
[3,]  614 15.39547 2.490118 17.307865  0.1889405
```

Figure 9. Medoid, Clustering vector, numerical information for each cluster

Figure 9 explains that the data selected as representative of the medoids are data 903, 1101, and 1446. Clustering is carried out on the 2.259 records displayed in the form of information on the number of data from each cluster. The maximum distance and the average value of the distance from each cluster. The number of data in cluster 1 is 1.024 data, with a maximum distance value of 7,93762 and an average distance value of 1,972092. The amount of data in cluster 2 is 1.621 with a maximum distance value of 11,96204 and an average distance value of 2,423424. The number of data in cluster 3 is 614 data, with a maximum distance value of 15,39547 and an average distance value of 2,490118. Figure 10 shows the results of the visualization of the K-Medoids algorithm, cluster 1 is marked in red, cluster 2 is marked in green, and cluster 3 is marked in blue.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

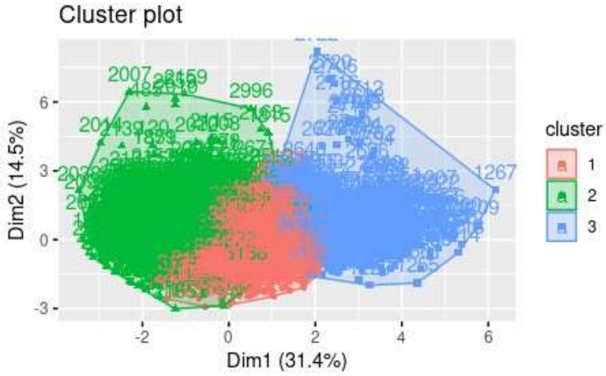
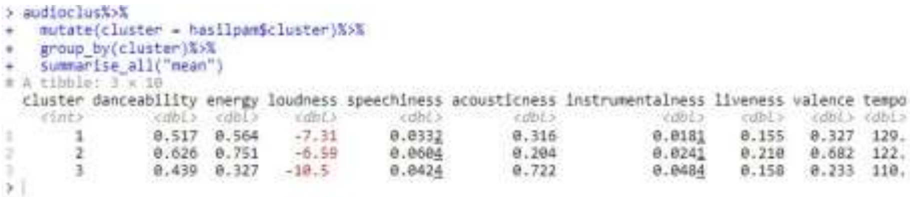


Figure 10. Visualization of K-Medoids algorithm

The final result of the K-Medoids clustering is the process of calculating the average of each cluster. Figure 11 shows the final result of the K-Medoids algorithm.



```
> audioclus%>%
+   mutate(cluster = hasilpam$cluster)%>%
+   group_by(cluster)%>%
+   summarise_all("mean")
# A tibble: 3 x 10
  cluster danceability energy loudness speechiness acousticness instrumentalness liveness valence tempo
    <int>        <dbl>  <dbl>    <dbl>       <dbl>        <dbl>            <dbl>    <dbl>   <dbl> <dbl>
1       1        0.517  0.564    -7.31      0.0332        0.316           0.0181    0.155   0.327  129.
2       2        0.626  0.751    -6.59      0.0604        0.204           0.0241    0.210   0.682  122.
3       3        0.439  0.327   -10.5       0.0424        0.722           0.0484    0.158   0.233  110.
> |
```

Figure 11. Final K-Medoids

Figure 11 explains that in cluster 1 it is Tempo because it has the highest value among other audio features reaching a value of 129, in cluster 2 it is Tempo because it has the highest value among other audio features reaching a value of 122, in cluster 3 it is Tempo because it has the highest value. High among the audio features reaches a value of 110.

3.5 The results of the comparison of the number of cluster members in the K-Means and K-Medoids algorithms.

After performing calculations on Rstudio, the number of cluster members for each attribute is obtained. Comparison results of K-Means and K-Medoids can be seen in table 2.

TABLE 2
Comparison Results Of K-Means And K- Medoids

Attributes and number of members	K-Means 1	K-Means 2	K-Means 3	K-Medoids 1	K-Medoids 2	K-Medoids 3
Danceability	0,481	0,607	0,516	0,517	0,626	0,439
Energy	0,422	0,736	0,614	0,564	0,751	0,327
Loudness	-9,06	-6,52	-8,90	-7,31	-6,59	-10,5
Speechiness	0,0381	0,0556	0,0398	0,0332	0,0604	0,0424
Acousticness	0,560	0,188	0,407	0,316	0,204	0,722
Instrumentalness	0,00733	0,00620	0,697	0,0181	0,0241	0,0484
Liveness	0,162	0,197	0,161	0,155	0,210	0,158
Valence	0,274	0,624	0,461	0,327	0,682	0,233
Tempo	118	125	123	129	122	110
Number of members	277	1208	1774	1024	1621	614

Based on table 2, it is explained that the highest number of cluster members in the K-Means algorithm occurs in cluster 3 with a total of 1.774 members, the number of medium cluster members occurs in cluster 2 with a total of 1.208 members, the lowest number of cluster members occurs in cluster 1 with a total of 277 members. While the highest number of cluster members in the K-Medoids algorithm occurs in cluster 2 with 1,621 members, the number of medium cluster members occurs in cluster 1 with a total of 1.024 members, and the lowest number of cluster members occurs in cluster 3 with a total of 614.

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

4. Conclusion

Based on the results of clustering audiotrack features based on the Indonesian pop genre by implementing data mining using Rstudio tools, the conclusions generated are as follows K Optimal on the K-Means and K-Medoids algorithms use the Silhouette Coefficient method using 3 clusters. The results obtained from the K-Means algorithm calculation process are based on the average, namely the Tempo audio feature because it has the highest value among other audio features. In cluster 1 it reached 118, in cluster 2 it reached 125, and in cluster 3 it reached 123. While the results obtained from the K-Medoids algorithm calculation process are based on the average, namely the Tempo audio feature because it has the highest value among other audio features. In cluster 1 it reached 129, in cluster 2 it reached 129, in cluster 3 reached 116. The results of the comparison of K-Means and K-Medoids, the highest cluster value for each member is found in the K-Means cluster 3 algorithm with a total of 1.774 members.

Reference

Otak, S. T., Neural, D., & View, N. (2021). Analisis Klaster Atribut Musik pada Global Top 50 dengan Data Spotify dengan Analisis Klaster Atribut Musik pada Global Top 50 dengan Data Spotify dengan Menggunakan Algoritma K-Means. January.

Putu, N., Merliana, E., & Santoso, A. J. (n.d.). Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means. 978–979.

Agustian, D. R., & Darmawan, B. A. (2022). Analisis Clustering Demam Berdarah Dengue Dengan Algoritma K-Medoids (Studi Kasus Kabupaten Karawang). JIKO (Jurnal Informatika Dan Komputer), 6(1), 18. https://doi.org/10.26798/jiko.v6i1.504

Issn, P., Zaidah, A. R., Septiarani, C. I., Nisa, S., & Wahyudi, N. (2021). Komparasi Algoritma K-Means , K-Medoid , Agglomeartive Clustering Terhadap Genre Spotify. 7(1), 49–54.

Athifaturrofifah, Goejantoro, R., & Yuniarti, D. (2019). Perbandingan Pengelompokan K-Means dan K-Medoids Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus : Data Titik Panas Di Indonesia Pada 28 April 2018). Jurnal EKSPONENSIAL, 10(2), 143–152.

Farahdinna, F., Nurdiansyah, I., Suryani, A., & Wibowo, A. (2019). Perbandingan Algoritma K-Means Dan K-Medoids Dalam Klasterisasi Produk Asuransi Perusahaan Nasional. Jurnal Ilmiah FIFO, 11(2), 208. https://doi.org/10.22441/fifo.2019.v11i2.010

Sahria, Y. (2020). Implementasi Teknik Web Scraping pada Jurnal SINTA Untuk Analisis Topik Penelitian Kesehatan Indonesia. URECOL (Unversity Research Colloqium), 297–306. http://repository.urecol.org/index.php/proceeding/article/view/1079

A. Yani, D. D., Pratiwi, H. S., & Muhardi, H. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. Jurnal Sistem Dan Teknologi Informasi (JUSTIN), 7(4), 257. https://doi.org/10.26418/justin.v7i4.30930

Sinambela, Y., Herman, S., Takwim, A., & Widianto, S. R. (2020). a Study of Comparing Conceptual and Performance of K-Means and Fuzzy C Means Algorithms (Clustering Method of Data Mining) of Consumer Segmentation. Jurnal Riset Informatika, 2(2), 49–54. https://doi.org/10.34288/jri.v2i2.116

Nabila, Z., Rahman Isnain, A., & Abidin, Z. (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

Herviany, M., Delima, S. P., Nurhidayah, T., & Kasini. (2021). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 1(1), 34–40.

Parlina, I., Windarto, A. P., Wanto, A., & Lubis, M. R. (2018). Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center. Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center Untuk Clustering Program Sdp, 3(1), 87–93.

Pramesti, D. F., Lahan, Tanzil Furqon, M., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 1(9), 723–732. https://doi.org/10.1109/EUMC.2008.4751704

Wira, B., Budianto, A. E., & Wiguna, A. S. (2019). Implementasi Metode K-Medoids Clustering Untuk

Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. RAINSTEK : Jurnal Terapan Sains & Teknologi, 1(3), 53–68. https://doi.org/10.21067/jtst.v1i3.3046

Suhanda, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. Jurnal Teknologi Informatika Dan Komputer, 6(2), 12–20. https://doi.org/10.37012/jtik.v6i2.299

Djufri, M. (2020). Penerapan Teknik Web Scraping Untuk Penggalian Potensi Pajak (Studi Kasus Pada Online Market Place Tokopedia, Shopee Dan Bukalapak). Jurnal BPPK : Badan Pendidikan Dan Pelatihan Keuangan, 13(2), 65–75. https://doi.org/10.48108/jurnalbppk.v13i2.636

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.