Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 5 No. 1 (2026) pp: 1865-1871
P-ISSN: 2963-9298, e-ISSN: 2963-914X

The Application of the K-Medoids Method for Clustering Meta Ads Audiences Based on Promotional Content Effectiveness

Ribut Julianto[1], Tedi Gunawan [2], Eko Aziz Apriadi[3]
[1,3]Informatics, Faculty of Science and Technology, University of Indonesia Mandiri
[2] Teknologi Informasi, Institut Teknologi Bisnis dan Bahasa Dian Cipta Cendikia
rjulian@uimandiri.ac.id[1] , tedigunawan73@gmail.com[2]

Abstract

The rapid growth of digital advertising platforms has encouraged businesses to adopt data-driven strategies in order to enhance the effectiveness of their promotional activities. One of the most widely used digital advertising services today is Meta Ads, which provides various performance metrics related to audience interactions. This study aims to segment Meta Ads audiences based on the effectiveness of promotional content using the K-Medoids clustering algorithm, which is known for its robustness in handling outliers compared to other clustering methods. The dataset used in this research consists of advertising access data obtained from ARTECH – PT. Arij Teknologi Inovasi. The data were processed and analyzed using RapidMiner as a data mining tool. After undergoing data preprocessing stages, including data cleaning and normalization, a total of 495 Meta Ads records were deemed suitable for clustering analysis. The results of the study show that the K-Medoids algorithm successfully grouped the data into two distinct clusters. Cluster 1 consists of 465 items and represents the dominant audience segment with relatively homogeneous interaction behavior, indicating consistent engagement patterns with promotional content. Meanwhile, Cluster 0 contains 30 items, representing a smaller but more specific audience segment with different access and interaction patterns. These findings demonstrate that the K-Medoids algorithm is effective in identifying meaningful audience segments from digital advertising data. The resulting clusters can be utilized to support more targeted digital marketing strategies, improve promotional content design, and optimize advertising budget allocation to achieve better campaign performance.

Keywords: *K-Medoids, Clustering, Meta Ads, Digital Marketing, Audience Segmentation, RapidMiner*

1. Introduction

The use of the internet has increasingly become an essential need for various age groups, including adolescents. Adolescence is a transitional period from childhood to adulthood, generally occurring between the ages of 12 and 21. During this stage, adolescents experience physical, social, and psychological changes and begin to engage in more egalitarian interactions with the adult community (Santrock, 2020). Recent data show that individuals aged 15–19 and 20–24 constitute the largest groups of internet users in Indonesia, confirming that adolescents are among the most intensive users of internet-based technologies (APJII, 2023).

Technological advancement has also driven the emergence of various communication innovations, one of which is the smartphone. The presence of smartphones enables users to access a wide range of social media features more easily and conveniently. Social media itself is a part of the internet that functions as a Web 2.0–based platform, allowing the creation and exchange of user-generated content. Various forms of social media, such as social networking sites and blogs, are now widely used by the public for communication and information acquisition in a fast, inexpensive, and easily accessible manner through mobile devices (Nasrullah, 2020).

Some of the most popular social media platforms among adolescents include Facebook and Instagram, which have become commonly installed and widely used applications on smartphones. Facebook remains widely utilized due to its comprehensive features, such as text messaging, photo and video sharing, stories, and integrated applications that support both social interaction and information exchange (Nasrullah, 2020). Similarly, Instagram attracts adolescent users through its visually oriented content and interactive features, enabling users to communicate, express themselves, and access information in a more engaging and efficient manner. Consequently, social media platforms such as Facebook and Instagram play a significant role in shaping adolescents' communication patterns and information consumption behaviors (Kominfo, 2022).

The growth of social media users in Indonesia has increased significantly in recent years. Reports from We Are Social and Meltwater indicate that Indonesia continues to be one of the countries with the largest number of social media users globally, with Facebook and Instagram consistently ranking among the most widely used platforms (Meltwater, 2023). Furthermore, official data from the Ministry of Communication and Informatics of the Republic of Indonesia (Kominfo) show that Facebook remains a dominant platform for digital advertising, with a substantial advertising audience reach. These platforms have become particularly popular among adolescents, who utilize social media not only for social interaction but also as a medium for searching information, products, and services relevant to their needs (Kominfo, 2022).

In the business context, marketing is a strategic activity that goes beyond mere selling or advertising. Marketing involves identifying consumer needs, developing products, and delivering value to customers in order to achieve sustainable business performance. Social media platforms, particularly Facebook and Instagram, have become important channels for product marketing due to their wide reach and interactive capabilities. Promotional activities on these platforms serve as a means to introduce products and communicate value through visual, verbal, and textual content (Siemens, 2005).

Along with technological advancements, traditional promotional media such as brochures, newspapers, television, and billboards have gradually declined in effectiveness, as they are considered less precise in targeting specific audiences. Digital media offer significant advantages by enabling advertisers to target audiences based on demographic and behavioral characteristics, including age, gender, interests, education, and online behavior. The widespread use of smartphones and social media has transformed the advertising industry, allowing advertising activities to be conducted more efficiently and at relatively lower costs. Facebook, in particular, provides advanced audience-targeting features that support more measurable and effective digital promotional strategies(Kominfo, 2022)(Dunleavy et al., 2011).

2. Research Methods

2.1. Data Mining

Data mining is a method used to process data with the objective of identifying hidden patterns within it. Through this process, the analyzed data can generate new information or knowledge derived from previously existing data, and these findings can be utilized as a basis for decision-making in the future (Julianto et al., 2025). In addition, data mining is also understood as a series of processes aimed at extracting valuable patterns from large-scale datasets, which can subsequently be stored in databases, data warehouses, or other information storage media (Han, J., Kamber, M., & Pei, 2022)

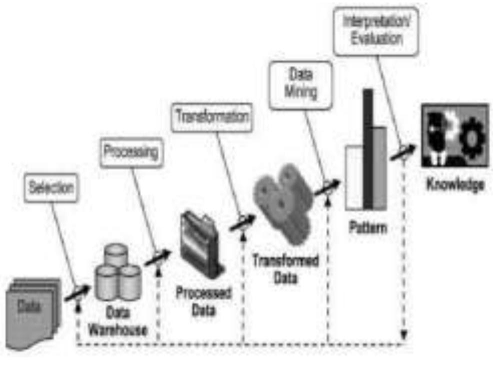

Figure 1. Stages of the Data Mining Process

2.2. Clustering

Clustering is a data analysis technique employed to partition a dataset into multiple clusters, such that objects within the same cluster exhibit a high degree of similarity, while objects belonging to different clusters demonstrate minimal similarity. This technique aims to discover inherent structures within data by grouping objects based on shared characteristics without requiring predefined class labels (Mustofa, Z., & Suasana, 2018).

In data mining and machine learning, clustering is classified as an unsupervised learning method, as it operates without prior knowledge of data categories. The grouping process is commonly performed by measuring the degree of similarity or dissimilarity between data objects using distance metrics such as Euclidean, Manhattan, or other similarity measures. The selection of an appropriate distance metric plays a crucial role in determining the quality and interpretability of the resulting clusters (Irianto et al., 2021).

Clustering techniques are widely applied in various domains, including market segmentation, pattern recognition, image processing, and social media analysis. In the context of digital marketing, clustering enables advertisers to identify distinct audience segments based on behavioral and interaction patterns, allowing for more targeted and effective promotional strategies. By understanding the characteristics of each cluster, organizations can optimize content delivery, improve customer engagement, and enhance decision-making processes (Amaluddin et al., 2015)

2.3. K-Medoids Algorithm

The K-medoids algorithm is a partition-based clustering technique that is closely related to the K-means and medoid-shift algorithms. It was originally introduced by Kaufman and Rousseeuw in 1987 under the name Partitioning Around Medoids (PAM). Similar to K-means, K-medoids aims to divide a dataset into a predefined number of clusters; however, it differs fundamentally in the manner in which cluster centers are determined. Instead of using the mean value of data points, K-medoids selects actual data objects as cluster representatives, referred to as medoids, which minimizes the total dissimilarity within each cluster (Mustofa, Z., & Suasana, 2018).

This distinctive characteristic makes K-medoids more robust to noise and outliers compared to K-means, particularly in datasets with skewed distributions or extreme values. Recent studies have highlighted that K-medoids is well suited for applications involving real-world data, where variability and irregular patterns are common, such as customer segmentation, digital marketing analytics, and social media data analysis (Mustofa, Z., & Suasana, 2018). Furthermore, the use of medoids enhances the interpretability of clustering results, as each cluster center corresponds to an actual observation within the dataset, which is advantageous for practical decision-making processes (Buaton, R., Sundari, Y., & Maulita, 2016).

In recent machine learning research, K-medoids continues to be applied and evaluated due to its balance between clustering accuracy and interpretability, especially when the reliability of cluster representation is prioritized over computational efficiency (Apriadi & Bisri, 2025). Consequently, K-medoids remains a relevant and effective clustering technique for exploratory data analysis in contemporary data-driven applications.

In the K-medoids algorithm, each cluster is represented by a medoid, which is an actual data object selected from the dataset. The medoid is chosen such that it minimizes the total dissimilarity between itself and all other data

$$d_{ij} = \sqrt{\sum_{a=1} (x_{ia} - x_{ja})^2} = \sqrt{(\mathrm{x}_i - \mathrm{x}_j)'(\mathrm{x}_i - \mathrm{x}_j)}$$

points within the same cluster. This characteristic makes K-medoids more robust to noise and outliers compared to K-means, which relies on mean values that can be significantly affected by extreme data points.

2.4. Data Analysis

The data analysis process was conducted after the collection of data and supporting evidence relevant to the objectives of the study. In this research, statistical data analysis, particularly descriptive statistics, was employed to examine, summarize, and describe the fundamental characteristics of the dataset. Descriptive statistical analysis plays an important role in providing an initial understanding of data distribution, central tendencies, and variability, which are essential for identifying patterns and anomalies prior to advanced analytical processes (Malik & Baharudin, 2013).

The data used in this study served as the primary foundation for supporting the implementation of the research methodology and subsequent analysis. To facilitate systematic processing and pattern discovery, the dataset was further analyzed using RapidMiner, a data mining and machine learning tool that supports preprocessing, clustering, and evaluation processes. The use of RapidMiner enables efficient handling of data transformation and analysis, thereby improving the reliability and accuracy of the results. Such tools are widely utilized in

DOI: https://doi.org/10.31004/riggs.v5i1.6515

contemporary data mining studies to enhance analytical consistency and support data-driven decision-making (Han, J., Kamber, M., & Pei, 2022).

2.5. Research Workflow Activity Diagram

The workflow carried out in this study is presented in the form of an activity diagram, as shown in Figure 2 below.

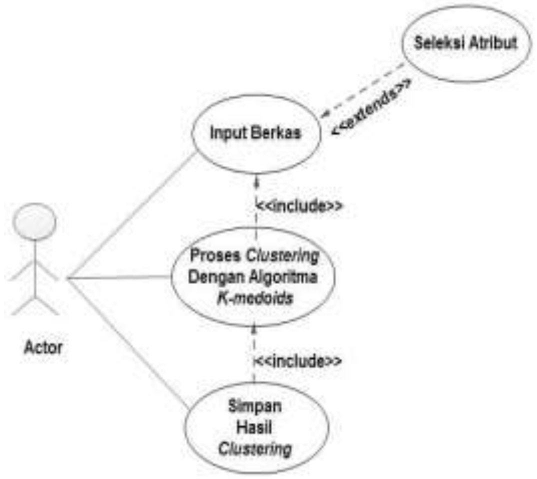

Figure 2. Research Workflow Activity Diagram

Figure 2 illustrates the workflow of activities carried out by the researcher within the implemented system. As shown in the figure, the researcher or user begins by identifying the research problem and defining the research objectives. Subsequently, data are collected from ARTECH – PT. Arij Teknologi Inovasi, which serve as the primary dataset for the study. The collected data are then validated using the performance evaluation features in the RapidMiner application. If the data are deemed valid, they are further processed using RapidMiner. Based on the information and results generated by the RapidMiner application, the researcher or user conducts an in-depth analysis of the processed data. Finally, decisions are made based on the research findings, and conclusions are drawn to summarize the outcomes of the study.

3. Results and Discussions

The data used in this study consist of product advertising data obtained from ARTECH – PT. Arij Teknologi Inovasi, collected over the period from March to October 2025. From an initial total of 1,114 records, a data preprocessing stage was conducted, resulting in 495 Meta Ads records that were properly organized and deemed suitable for use as the dataset for further analysis. The detailed distribution of the data is presented in the following table.

Table 1. Meta Ads Advertising Access Data

No	Ads Code	Device	Gender	Number of Accesses
1	PURCH10	1	1	2
2	PURCH10	1	1	2
3	PURCH10	1	1	4
4	PURCH10	1	1	1
5	PURCH10	1	2	2
6	PURCH10	1	1	3
7	PURCH11	1	1	2
8	PURCH09	1	1	1
9	PURCH10	1	1	1
10	PURCH10	1	1	3
11	PURCH09	1	2	1
12	PURCH10	1	1	1
13	PURCH10	1	1	2

DOI: https://doi.org/10.31004/riggs.v5i1.6515

14	PURCH10	1	1	1
15	PURCH11	1	1	2
16	PURCH09	1	1	1
17	PURCH10	1	1	4
18	PURCH10	1	1	2
19	PURCH09	1	1	2
20	PURCH10	1	1	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
495	PURCH10	1	1	1
	MIN	1	1	1
	MAX	2	2	54

3.1. K-Medoids Data Processing

The determination of the number of clusters (k) for the n objects in this study resulted in two clusters.

1. The initial centroids (medoids) were assumed and selected randomly at the beginning of the clustering process. In this study, the selected initial centroids are presented in the following table.

Table 2 Medoids Meta Ads

No	Device	Gender	Number of Accesses
100	0,000	1,000	0,245
101	0,000	0,000	0,019

2. The medoid objects were assigned to the nearest clusters based on the Euclidean distance measure. The following section presents the distance calculations for the promotional case data based on advertising metrics and the number of visits.

$$N1 = \sqrt{(0{,}000 - 0{,}000)^2 + (1{,}000 - 0{,}000)^2 + (0{,}245 - 0{,}019)^2} = 1{,}025$$

$$N2 = \sqrt{(0{,}000 - 0{,}000)^2 + (0{,}000 - 0{,}000)^2 + (0{,}019 - 0{,}019)^2} = 0{,}000$$

3.2 K-Medoids Implementation Using RapidMiner

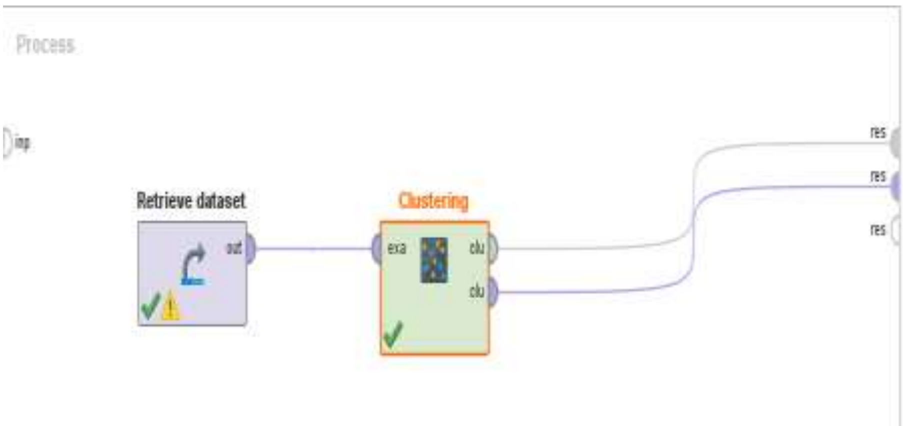

Figure 3. K-Medoids Implementation Using RapidMiner

DOI: https://doi.org/10.31004/riggs.v5i1.6515

Cluster Model

```
Cluster 0: 30 items
Cluster 1: 465 items
Total number of items: 495
```

Figure 4. K-medoids parameter

Based on the results of the clustering implementation using RapidMiner, the K-Medoids algorithm successfully segmented the Meta Ads dataset into two distinct clusters. As illustrated in the cluster model output, Cluster 0 consists of 30 items, while Cluster 1 contains 465 items, from a total of 495 data items. This distribution indicates an imbalanced clustering outcome, where the majority of advertising data points are concentrated in Cluster 1. The dominance of Cluster 1 suggests that most advertising records share similar characteristics in terms of device usage, gender category, and the number of ad accesses. This cluster can be interpreted as representing the general or common audience segment, characterized by relatively homogeneous interaction patterns and lower variability in promotional response metrics. The high number of items in this cluster reflects a consistent behavior pattern among the majority of users exposed to the advertisements.

In contrast, Cluster 0, which contains a significantly smaller number of items, represents a more distinct and specific audience segment. The limited size of this cluster indicates that the data points within it possess characteristics that differ noticeably from the dominant pattern observed in Cluster 1. This cluster may correspond to audiences with unique interaction behaviors, such as higher or lower access frequencies, different gender responses, or specific device usage patterns, making them stand out during the distance-based clustering process. The assignment of data objects to clusters was performed based on Euclidean distance, ensuring that each data point was grouped with the medoid to which it had the smallest dissimilarity. The use of K-Medoids, which selects actual data points as cluster centers, enhances the robustness of the clustering results, particularly in handling potential outliers in advertising access data. This characteristic is especially relevant in digital marketing datasets, where user interaction metrics often exhibit uneven distributions.

Overall, the clustering results provide meaningful insights into audience segmentation for Meta Ads. The identification of a dominant cluster and a smaller, distinct cluster enables advertisers to differentiate between mainstream audiences and specialized audience segments. Such segmentation can support more targeted advertising strategies, allowing marketers to optimize content design, audience targeting, and budget allocation based on the behavioral characteristics of each cluster.

4. Conclusion

This study successfully implemented a machine learning approach using the K-Medoids clustering algorithm to segment Meta Ads audiences based on the effectiveness of promotional content. The results demonstrate that K-Medoids is capable of identifying meaningful audience segments from unlabeled advertising data by grouping records with similar interaction characteristics. Using RapidMiner, a total of 495 Meta Ads records were clustered into two groups, with Cluster 1 comprising the majority of the data (465 items) and Cluster 0 representing a smaller but distinct segment (30 items). This clustering outcome indicates the presence of a dominant audience group with relatively homogeneous behavior, alongside a more specific segment exhibiting different interaction patterns.The findings provide valuable insights for digital marketing strategies, as the resulting audience segments can be utilized to improve targeting accuracy, optimize promotional content, and allocate advertising budgets more effectively. By distinguishing between dominant and specialized audience groups, advertisers can design more tailored and data-driven campaigns. Future research is recommended to incorporate larger datasets, additional behavioral and demographic variables, and comparative analyses with other clustering algorithms to further enhance the robustness and applicability of audience segmentation models.

Reference

1. Amaluddin, F., Muslim, M., & Naba, A. (2015). Klasifikasi Kendaraan Menggunakan Gaussian Mixture Model (GMM) Dan Fuzzy Cluster K Means (FCM). *Jurnal EECCIS*, *9*(1), pp.19-24.
2. APJII. (2023). Laporan Survei Internet APJII 2023–2024. *Asosiasi Penyelenggara Jasa Internet Indonesia*.
3. Apriadi, E. A., & Bisri, M. (2025). Optimization of BPJS Health Facility Distribution with K-Means Clustering Algorithm.

International Journal of Technology and Computer Science, *1*(1), 1–13.
4. Buaton, R., Sundari, Y., & Maulita, Y. (2016). Clustering Tindak Kekerasan Pada Anak Menggunakan Algoritma K-Means Dengan Perbandingan Jarak Kedekatan Manhattan City Dan Euclidean. *Media Informasi Analisa Dan Sistem, 1 (2)*, 47–53.
5. Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2011). Digital Era Governance: IT Corporations, the State, and e-Government. In *Digital Era Governance: IT Corporations, the State, and e-Government*. https://doi.org/10.1093/acprof:oso/9780199296194.001.0001
6. Han, J., Kamber, M., & Pei, J. (2022). *Data Mining: Concepts and Techniques (4th ed.)*. Morgan Kaufmann.
7. Irianto, S. Y., Yulianto, R., Karnila, S., & Yuliawati, D. (2021). Studi Akurasi Karakteristik Retina sebagai Future Identification dengan Euclidean Distance Metrics. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, *16*(1), 19. https://doi.org/10.30872/jim.v16i1.5136
8. J. W. Santrock. (2020). Adolescence (17th ed., Indonesian context reference). *McGraw-Hill Education*.
9. Julianto, R., Gunawan, T., & Apriadi, E. A. (2025). *Development of an Intelligent Educational Chatbot Using NLP and Machine Learning*. *1*(1), 14–24.
10. Kominfo. (2022). *Perkembangan Ekonomi Digital dan Media Sosial di Indonesia*. Kementerian Komunikasi dan Informatika Republik Indonesia.
11. Malik, F., & Baharudin, B. (2013). Analysis of distance metrics in content-based image retrieval using statistical quantized histogram texture features in the DCT domain. *Journal of King Saud University - Computer and Information Sciences*, *25*(2), 207–218. https://doi.org/10.1016/j.jksuci.2012.11.004
12. Meltwater, W. A. S. &. (2023). *Digital 2023*.
13. Mustofa, Z., & Suasana, I. S. (2018). Algoritma Clustering K-Medoids Pada E-Government Bidang Information And Communication Technology Dalam Penentuan Status Edgi. *Jurnal Teknologi Informasi Dan Komunikasi*, *9 (1)*.
14. Nasrullah, R. (2020). *Media Sosial: Perspektif Komunikasi, Budaya, dan Sosioteknologi (Edisi Revisi)*. Simbiosa Rekatama Media.
15. Siemens, G. (2005). Connectivism: A Learning Theory for the Digital Age. *International Journal of Instructional Technology and Distance Learning*.

DOI: https://doi.org/10.31004/riggs.v5i1.6515