

Comparison Analysis of K-Means and DBSCAN Algorithms for Improving Budget Absorption Efficiency in EIS

Fery Salman Farisy^{1*}, Fauziah²

^{1,2}Universitas Nasional, Indonesia

¹ferysalman2022@student.unas.ac.id, ²fauziah@student.unas.ac.id



ABSTRACT

This study aims to analyze the comparison between the K-Means and DBSCAN clustering algorithms in budget absorption within the Executive Information System (EIS). Realized budget achievement data from regional devices serve as the primary dataset for analysis. Before conducting experiments, the data undergo a preprocessing stage to eliminate outliers and apply normalization processes, ensuring the data is ready for further analysis. Subsequently, both algorithms, K-Means and DBSCAN, are applied to the budget achievement data to generate clusters corresponding to their respective characteristics. The research anticipates that the results will unveil significant findings in comparing the performance of K-Means and DBSCAN algorithms in budget absorption within the EIS context, especially within the scope of this study. Therefore, this analysis is expected to provide valuable insights for stakeholders aiming to enhance the efficiency and effectiveness of budget management through the optimal utilization of clustering algorithms within the EIS.

*Corresponding Author

Article History:

Submitted: 07-12-2023

Accepted: 20-12-2023

Published: 22-12-2023

Keywords:

EIS; DBSCAN; K-Means

Brilliance: Research of Artificial Intelligence is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

INTRODUCTION

In the era of technological advancements, Executive Information Systems (EIS) have become crucial components in budget management within regional devices ('SEGMENTASI SATUAN KERJA DALAM RANGKA AKSELERASI PENYERAPAN ANGGARAN DAN MENDORONG EKONOMI REGIONAL', 2023). It enables decision-makers to monitor and manage budget realization achievements effectively (Bratakusuma & Rifai, 2019). In this context, this research aims to conduct a comparative analysis between two clustering algorithms, namely K-Means and DBSCAN, with the goal of improving budget absorption efficiency in EIS. K-Means and DBSCAN algorithms are commonly used clustering methods to group data based on similar characteristics and significant differences (Bestari Ing Sukrisno Mardiyanto Anwar Soekito Ardjo, Hadi Sutopo, & Ketua Penyunting Sularso Budilaksono, 2018). By conducting a comparative analysis, the researchers will provide guidance in selecting the algorithm that is more suitable for improving budget absorption efficiency within EIS (Bratakusuma & Rifai, 2019). For this research, the realized budget achievement data from each regional device are used as the primary data for analysis. Before conducting the experiments, the data undergo a preprocessing stage to remove outliers and apply normalization processes to ensure that the data is ready for further analysis (Murti Suyoto, Rachmadi, & Parulian, 2022). Afterward, both algorithms, K-Means and DBSCAN, are applied to the realized budget achievement data to generate clusters that correspond to their respective characteristics.

It is expected that this research will uncover significant findings in comparing the performance between the K-Means and DBSCAN algorithms in budget absorption within EIS, particularly in the context of this study. Therefore, this analysis is expected to provide valuable insights for stakeholders in their efforts to improve the efficiency and effectiveness of budget management through the optimal utilization of clustering algorithms within EIS.

LITERATURE REVIEW

Bratakusuma and Rifai (2019) delved into the role of Executive Information Systems (EIS) in supporting budgeting in village governance through the integration of systems with web service technology and mobile applications. Focused on the village of Dermaji, the study aimed to comprehend how EIS could serve as a pillar in budget management at the village level. Furthermore, Murti Suyoto, Rachmadi, and Parulian (2022) highlighted the application of the K-Means algorithm to determine appropriate clusters in measuring the effectiveness of budget implementation in the Ministry of Agrarian and Spatial Planning/National Land Agency. This study contributed to understanding how data clustering with K-Means could assist in evaluating budget implementation within governmental institutions.

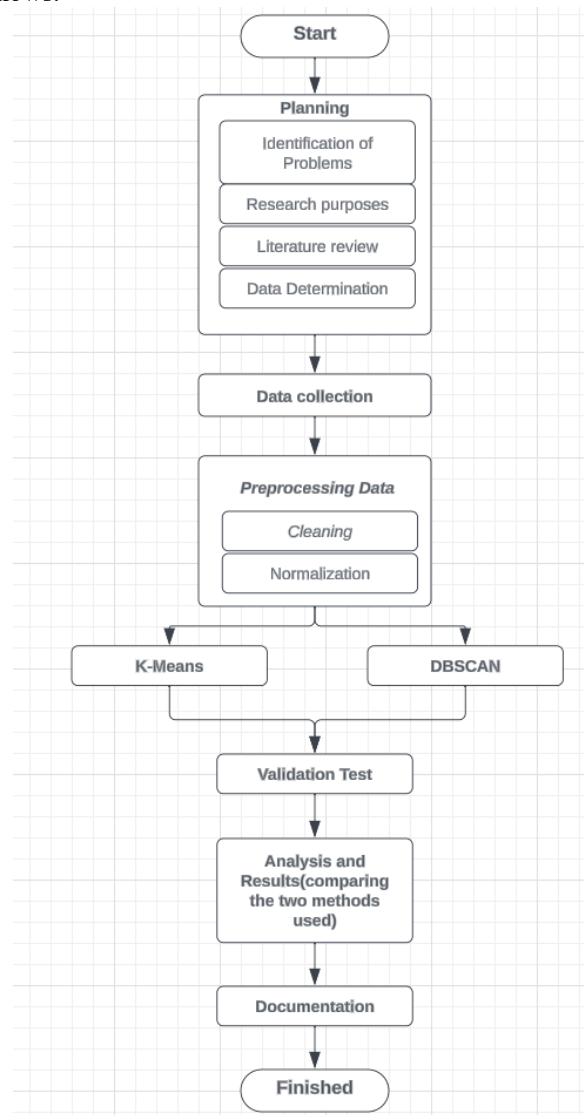
On the other hand, the research by Darmawan, Jenderal, Keuangan, Purnomo, Nilawati, and Perdiansyah (2022) placed a stronger emphasis on implementing Machine Learning algorithms to develop a predictive model for the Non-Performing Loan (NPL) rate, with the goal of credit guarantee for Micro, Small, and Medium Enterprises (MSMEs) as



part of national economic recovery. They explored how this algorithm could be a crucial instrument in anticipating and managing credit risks. In line with this, another study by Murti Suyoto, Rachmadi, and Parulian (2022) remained within the context of using the K-Means algorithm, this time to measure the effectiveness of budget implementation in the Ministry of Agrarian and Spatial Planning/National Land Agency. Overall, the findings from these studies synergize to enrich the understanding of the role of clustering algorithms, particularly K-Means, in governmental sector budget management.

METHOD

The research method begins with a planning phase, focusing on comparing the K-Means and DBSCAN algorithms in budget absorption clustering. The data source used is the local government budget for one district in the fiscal year 2022, obtained directly from the relevant authorities. The initial dataset undergoes a cleaning process to remove invalid entries. Subsequently, the data is normalized to be processed by both algorithms, K-Means and DBSCAN (Benri, Metisen, & Latipa, 2015; Jing, Zhao, & Jiang, 2019). The research method can be further explained through reference to Figure 1 as follows:



Picture 1. Research methodology

K-Keans

K-Means is one of the methods that can group data into two or more clusters (*PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING DATA ANGGARAN PENDAPATAN BELANJA DAERAH DI KABUPATEN XYZ SKRIPSI*, n.d.). This method clusters information in such a way that information with similar qualities is gathered in similar clusters, while information with different attributes is placed in different clusters (Mayang Sari, n.d.). K-Means is the simplest algorithm and the most commonly used clustering algorithm (Bestari Ing Sukrisno Mardiyanto Anwar

Soekito Ardjo et al., 2018; Monalisa & Kurnia, 2019). In calculating the distance of the i -th data point (x_i) to the k -th cluster (c_k), the Euclidean formula can be used with an equation like Equation 1:

$$d_{ik} = \sqrt{\sum_{j=1}^m (c_{ij} - x_{ik})^2} \quad (1)$$

Density-Based Spatial Clustering of Applications with noise (DBSCAN)

The DBSCAN algorithm can identify core samples with high density and expand clusters from the cluster-determining algorithm. It requires a minimum number of samples and the epsilon (ϵ) parameter, which represents the maximum distance between samples to be considered part of the same cluster (Monalisa, Juniarti, Saputra, Muttakin, & Ahsyar, 2023; Mu, Hou, Zhao, Wei, & Wu, 2023). The main limitation is determining the number of basic features that need to be characterized together. The algorithm's tolerance for noise is determined by this parameter.

- 1) Determine the values for MinPts and Eps.
- 2) Choose an initial point or value, p , randomly.
- 3) Calculate Eps using the Euclidean distance formula for all points that are density reachable from p .

$$d_{ij} = \sqrt{\sum_a^p (x_{ia} - x_{ja})^2} \quad (2)$$

Where x_{ia} is the a -th variable of object i ($i = 1, \dots, n$; $a = 1, \dots, p$) and d_{ij} is the Euclidean distance value between two objects i and j .

- 4) A cluster is formed when the number of points that satisfy the Eps condition is greater than the value of MinPts, and the point p is a core point.
- 5) Repeat steps 3-4 until all focal points are handled. If P is a point on a line and there are no focal points reachable with the thickness reachable of P , then the cycle continues to another point.

RESULT

This research aims to analyze budget absorption in Executive Information Systems (EIS) based on regional devices using a clustering approach with the K-Means and DBSCAN algorithms. The selection of these two algorithms aims to gain holistic insights into budget absorption patterns and to understand resilience to variations and outliers.

Data Collection and Preprocessing

The budget data for regional devices is obtained from official financial sources and includes monthly expenditure details. The information in the dataset comprises variables such as regional device, account, budget, actual spending, and month.

tahun	bulan	perangkat_daerah	anggaran	realisasi
2022	1	Dinas Pendidikan	621902373968	19643859279
2022	2	Dinas Pendidikan	621902373968	19843236476
2022	3	Dinas Pendidikan	621902373968	19525270478
2022	4	Dinas Pendidikan	621902373968	77199619745
2022	5	Dinas Pendidikan	621902373968	22795986394
2022	6	Dinas Pendidikan	621902373968	51004240129
2022	7	Dinas Pendidikan	621902373968	86531112068
2022	8	Dinas Pendidikan	621902373968	31542797700
2022	9	Dinas Pendidikan	621902373968	61012657196
2022	10	Dinas Pendidikan	621902373968	28746250161

Picture 2. Data Collection

Normalization Data

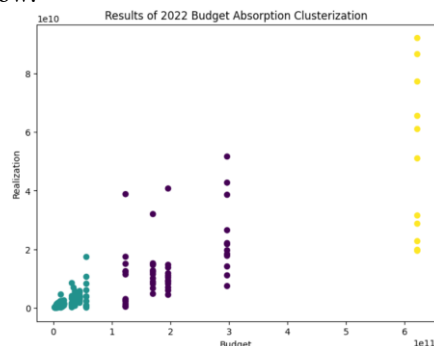
The normalization applied in this research is min-max scaling. The normalized data results can be seen in Picture 2.

tahun	bulan	perangkat_daerah	anggaran	realisasi
2022	1	Dinas Pendidikan	621902373968	0.212612
2022	2	Dinas Pendidikan	621902373968	0.214781
2022	3	Dinas Pendidikan	621902373968	0.211322
2022	4	Dinas Pendidikan	621902373968	0.838697
2022	5	Dinas Pendidikan	621902373968	0.246900
2022	6	Dinas Pendidikan	621902373968	0.553747
2022	7	Dinas Pendidikan	621902373968	0.940204
2022	8	Dinas Pendidikan	621902373968	0.342047
2022	9	Dinas Pendidikan	621902373968	0.662617
2022	10	Dinas Pendidikan	621902373968	0.311627

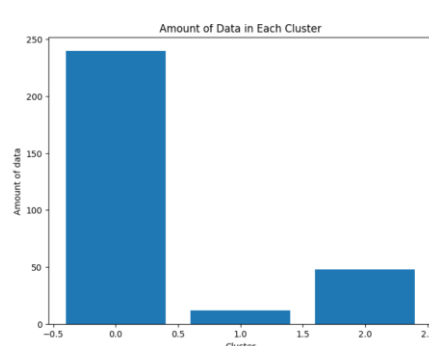
Picture 3. Normalization Data

Clustering with the K-Means Algorithm

In the clustering analysis conducted, the K-Means algorithm is applied to the budget absorption dataset for the year 2022. This dataset consists of columns for budget and actual spending, which are used as features for clustering. The use of the scikit-learn library allows for the implementation of the K-Means algorithm. In the clustering process, the desired number of clusters is three. After the clustering is complete, cluster labels are added to the dataset, indicating the membership of each data point in the corresponding cluster. The clustering results are displayed in a table that shows the budget value, actual spending value, and cluster label for several initial data points (Benabdellah, Benghabrit, & Bouhaddou, 2019). This enables the observation of patterns in data grouping based on similar budget and actual spending characteristics within each cluster. The results of clustering the data using the K-Means algorithm can be seen in Figures 4 and 5 below.



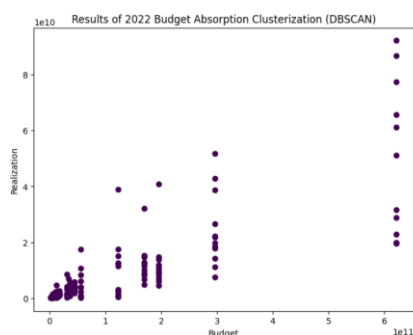
Picture 4. Budget bsorption Clusterization K-Means



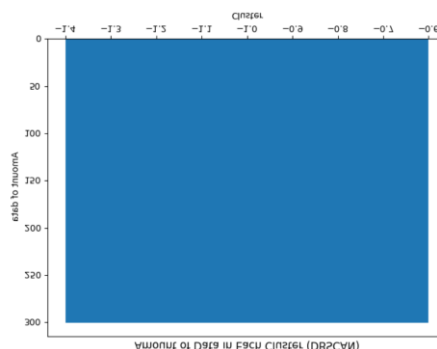
Picture 5. Amount of Data in Each Cluster K-Means

Clustering with the DBSCAN Algorithm

In cluster analysis, this time the DBSCAN algorithm was used on the 2022 budget absorption dataset. The dataset consists of budget and actualization columns, which are the main features in the clustering process. By using the scikit-learn library, the DBSCAN algorithm is applied to group data based on the proximity distance between samples. In the code, the parameter "eps" sets the maximum distance between two samples to be classified into the same cluster, while "min_samples" determines the minimum number of samples within a certain radius to form a cluster. After the clustering process is completed, cluster labels are added to the dataset, showing the membership of each data in the corresponding cluster. To visualize the clustering results, a scatter plot is used with the x and y axes representing the budget and actualization values. Each point on the scatter plot is marked with a different color based on the cluster it belongs to. By observing this scatter plot, you can observe patterns of data grouping based on similar budget and actualization characteristics within each cluster. The results of clustering the data using the DBSCAN algorithm can be seen in Figures 6 and 7 below.



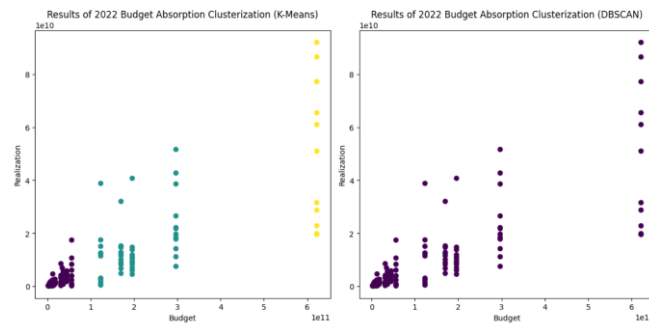
Picture 6. Budget bsorption Clusterization DBSCAN



Picture 7. Amount of Data in Each Cluster DBSCAN

Comparison of K-Means and DBSCAN Algorithms

The conclusion is that the K-Means algorithm can significantly improve budget absorption efficiency in the SIE. The K-Means algorithm has the advantage of producing quality clusters that help understand patterns and trends in budget actualization achievements (Awaad & Hefny, n.d.; Darmawan et al., 2022). On the other hand, DBSCAN is more suitable to use when there are irregular or complex-shaped clusters. The results of comparing data clustering using the K-Means and DBSCAN algorithms can be seen in Figure 8 below.



Picture 8. Budget bsorption Clusterization K-Means and DBSCAN

DISCUSSION

In this chapter, we will discuss and analyze in-depth the findings obtained from the research on the comparative analysis between the K-Means and DBSCAN algorithms in budget absorption within the Executive Information System (EIS). This discussion aims to provide a deeper interpretation of these findings and to broaden understanding of the implications and significance of this research for budget processing within the EIS.

Implications of Findings

From this research, it was found that the use of the K-Means algorithm can significantly improve budget absorption efficiency within the EIS. The advantage of the K-Means algorithm in producing high-quality clusters brings important benefits in understanding patterns and trends in budget actualization achievements. On the other hand, DBSCAN is a more suitable choice when dealing with groups that exhibit irregular or complex cluster characteristics.

Relevance to Budget Management in SIE

The application of clustering algorithms in the EIS has significant relevance in achieving more effective budget grouping. The optimal use of clustering algorithms can result in groups with similar characteristics, which in turn facilitates better analysis and decision-making in budget management. By selecting the appropriate clustering algorithm based on the needs and characteristics of the data, the efficiency of budget absorption in the EIS can be improved.

CONCLUSION

Based on the research on the comparative analysis between the K-Means and DBSCAN algorithms in budget absorption within the Executive Information System (EIS), it was found that the use of the K-Means algorithm can improve budget absorption efficiency in the EIS by producing high-quality clusters. DBSCAN is more suitable for irregular or complex groups. The optimal application of clustering algorithms in the EIS can facilitate better analysis and decision-making in budget management. This research has limitations in terms of the selection of clustering algorithms, so future research is recommended to consider other algorithms and other factors that influence budget absorption in the EIS. In conclusion, this study provides important guidance in improving budget absorption efficiency through the appropriate utilization of clustering algorithms in the EIS.

ACKNOWLEDGMENT

I would like to express my gratitude to all parties who have provided support and contribution to this research. I also want to thank my institution for providing the necessary facilities and resources for this study. I would like to express my appreciation to the respondents who participated in this research by providing budget actualization data from local government agencies. I am also grateful to my family for their support and understanding throughout the course of this research. Thank you for all the support and valuable contributions from everyone mentioned. Without their support and contributions, this research would not have been possible.

REFERENCES

- Awaad, A. M., & Hefny, H. (n.d.). Parallel Implementation of Statistical DBSCAN Algorithm for Spark-based Clustering on Google Cloud Platform. *International Journal of Computing and Digital Systems*, 20, 2210–142. doi:10.12785/ijcds/XXXXXX
- Benabdellah, A. C., Benghabrit, A., & Bouhaddou, I. (2019). *A survey of clustering algorithms for an industrial context*. In *Procedia Computer Science* (Vol. 148, pp. 291–302). Elsevier B.V. doi:10.1016/j.procs.2019.01.022
- Benri, M., Metisen, H., & Latipa, S. (2015). *ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK PADA SWALAYAN FADHILA*. *Jurnal Media Infotama* (Vol. 11).

- Bestari Ing Sukrisno Mardiyanto Anwar Soekito Ardjo, M. M., Hadi Sutopo, Mk., & Ketua Penyunting Sularso Budilaksono, M. (2018). *DEWAN REDAKSI IKRA-ITH INFORMATIKA* (Vol. 2). Retrieved from www.journals.upi-yai.ac.id
- Bratakusuma, T., & Rifai, Z. (2019). SISTEM INFORMASI EKSEKUTIF SEBAGAI PENDUKUNG PENGANGGARAN DI PEMERINTAH DESA MELALUI INTEGRASI SISTEM DENGAN TEKNOLOGI WEB SERVICE DAN APLIKASI SELULER (STUDI KASUS DESA DERMAJI) EXECUTIVE INFORMATION SYSTEMS AS A SUPPORT OF BUDGETING IN A VILL GOVERNMENT THROUGH AND CELLULAR APPLICATION (CASE STUDY OF DERMAJI VILL), 2(2), 9–16.
- Darmawan, R., Jenderal, I., Keuangan, K., Purnomo, D., Nilawati, S., & Perdiansyah, F. (2022). *Implementasi Algoritma Machine Learning untuk Pengembangan Model Prediktif atas Tingkat Non-Performing Loan dalam rangka Penjaminan Kredit UMKM Pemulihan Ekonomi Nasional* (Vol. 1).
- Jing, W., Zhao, C., & Jiang, C. (2019). *An improvement method of DBSCAN algorithm on cloud computing*. In *Procedia Computer Science* (Vol. 147, pp. 596–604). Elsevier B.V. doi:10.1016/j.procs.2019.01.208
- Mayang Sari, R. (n.d.). *Prediksi Data Anggaran Pendapatan Belanja Daerah Menggunakan Algoritma K-Means*.
- Monalisa, S., Juniarti, Y., Saputra, E., Muttakin, F., & Ahsyar, T. K. (2023). Customer segmentation with RFM models and demographic variable using DBSCAN algorithm. *Telkomnika (Telecommunication Computing Electronics and Control)*, 21(4), 742–749. doi:10.12928/TELKOMNIKA.v21i4.22759
- Monalisa, S., & Kurnia, F. (2019). Analysis of DBSCAN and K-means algorithm for evaluating outlier on RFM model of customer behaviour. *Telkomnika (Telecommunication Computing Electronics and Control)*, 17(1), 110–117. doi:10.12928/TELKOMNIKA.v17i1.9394
- Mu, C., Hou, Y., Zhao, J., Wei, S., & Wu, Y. (2023). Stream-DBSCAN: A Streaming Distributed Clustering Model for Water Quality Monitoring. *Applied Sciences (Switzerland)*, 13(9). doi:10.3390/app13095408
- Murti Suyoto, I. D., Rachmadi, T., & Parulian, L. T. (2022). MENENTUKAN CLUSTER YANG TEPAT DENGAN K-MEANS DALAM RANGKA MENGUKUR EFEKTIVITAS PELAKSANAAN ANGGARAN PADA KEMENTERIAN AGRARIA DAN TATA RUANG/BADAN PERTANAHAN. *Infotech: Journal of Technology Information*, 8(1), 13–22. doi:10.37365/jti.v8i1.126
- PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING DATA ANGGARAN PENDAPATAN BELANJA DAERAH DI KABUPATEN XYZ SKRIPSI. (n.d.).
- SEGMENTASI SATUAN KERJA DALAM RANGKA AKSELERASI PENYERAPAN ANGGARAN DAN MENDORONG EKONOMI REGIONAL. (2023). *Jurnal Manajemen Perbendaharaan*, 4(1), 73–93. doi:10.33105/jmp.v4i1.487