Proceeding of International Conference Technology, Economics, and Social Science
Vol 1, No 1, January 2026, Page 749-757
ISSN 3124-6087 (Media Online)
https://journals.adaresearch.or.id/ictess

Development of an Intelligent Hybrid Clustering Algorithm Based on Depth Data and Grid Mapping for Automatic Parameter Estimation

Sutrisno, Sandi Ardiansyah, Hartono*, Sayuti Rahman

Postgraduate Program, Magister of Informatics Study Program, Universitas Medan Area, Medan, Indonesia
Jl. Setia Budi No. 79 B / Jalan Sei Serayu No. 70 A, Medan 20112, Sumatera Utara, Indonesia
Email: [1]sutrisno21@gmail.com, [2]ardiansyah.sandi06@gmail.com, [3,*]hartono@staff.uma.ac.id, [4]sayutirahman@staff.uma.ac.id
Correspondence Author Email: hartono@staff.uma.ac.id
Submitted: **01/12/2025**; Accepted: **07/01/2026**; Published: **25/01/2026**

Abstract−The K-Means algorithm is one of the most popular clustering methods; however, it has notable weaknesses, particularly its reliance on an arbitrary selection of the number of clusters (K) and its sensitivity to the initialization of centroid positions. Traditional K-Means often exhibits unstable performance and is vulnerable to noise, while many cluster validity indices (CVIs) used to determine the optimal K also suffer from limitations related to computational complexity and efficiency. Although advanced approaches such as grid-based clustering and Data Depth–based estimation methods have been developed to address these issues, many of them still require manual intervention to determine key parameters, which becomes an obstacle in creating a fully autonomous clustering system. To overcome these challenges, this study proposes the development of an Intelligent Hybrid Clustering Algorithm that integrates Grid-Mapping with the concept of Data Depth for automatic parameter estimation. The Grid-Mapping approach is employed due to its proven speed, stability, and robustness against noise by transforming data into a grid-based representation. Meanwhile, the Data Depth concept is utilized as a foundation for efficiently and accurately estimating the optimal number of clusters (K) without the need for repeated full clustering processes. The main innovation of this research lies in creating an automated mechanism for determining crucial parameters, thus eliminating the need for manual user input. With the integration of these two approaches, the proposed algorithm is expected to offer a clustering solution that is not only accurate and efficient but also more intelligent and adaptive, capable of operating autonomously across various data scenarios.

Keywords: Clustering, K-Means, Hybrid Algorithm, Data Depth, Grid-Mapping, Automatic Parameter Estimation

1. INTRODUCTION

In the era of big data, unsupervised learning algorithms such as clustering have become increasingly prominent (Guo et al., 2024). Clustering is a data mining technique used to group objects based on similarities in their characteristics (Utami, 2023), with the primary objective of maximizing intra-cluster similarity while minimizing inter-cluster similarity (Guo et al., 2024). Its ability to identify natural patterns and groups within data makes it a fundamental component of many modern applications (Pugliese et al., 2021). However, the effectiveness of many clustering algorithms particularly K-Means is often hindered by several significant challenges. One of the most widely used clustering algorithms is K-Means, known for its simplicity and effectiveness (Pugazhenthi & Kumar, 2020). Nevertheless, the traditional K-Means algorithm faces key limitations, especially in the arbitrary selection of the number of clusters (K), which often results in trial-and-error procedures to determine the Optimal Number of Clusters (ONC) (I. K. Khan et al., 2024). In addition, the performance of K-Means is highly sensitive to the selection of initial random centroids, which can significantly influence the segmentation results and increase computational complexity (Pugazhenthi & Kumar, 2020). To address the challenge of determining the optimal number of clusters, various methodologies have been proposed, one of which is the use of Cluster Validity Indices (CVIs) (I. K. Khan et al., 2024). Popular indices include the Elbow method (Prastyabudi et al., 2024) and the Silhouette Coefficient (Utami, 2023). However, each method has its own limitations. The Gap Statistic method, for instance, depends on expected values derived from reference data, posing constraints when dealing with datasets of varying scales and overlapping distributions (I. K. Khan et al., 2024). Meanwhile, the Elbow method often struggles with complex data distributions, potentially resulting in ambiguous or misleading ONC determination (I. K. Khan et al., 2024).

Alongside these approaches, more advanced techniques such as density-based and grid-based methods have been developed to address the shortcomings of partition-based algorithms like K-Means (Tareq et al., 2022). Density-grid-based clustering adopts a grid data structure in which data objects are mapped into grid cells, and clusters are formed based on cell density (Tareq et al., 2022). This approach is advantageous for detecting arbitrarily shaped clusters and filtering noise or outliers (Tareq et al., 2022). To further enhance clustering performance, hybrid clustering algorithms that combine multiple techniques have been introduced (Guo et al., 2024). One example is the integration of Density Peak Clustering (DPC) with Mean-Shift to address DPC's sensitivity to parameters and improve the accuracy of cluster center identification (Guo et al., 2024). Despite these advancements, several research gaps remain. Existing grid-based algorithms often produce low-quality clusters for continuously evolving data or data streams (Tareq et al., 2022). Many of these methods also still require manual parameter tuning. This dependence on manual intervention becomes a major obstacle in developing fully autonomous and intelligent clustering systems. Based on these gaps, this study aims to develop an intelligent hybrid clustering algorithm capable of performing automatic parameter estimation. This research integrates the strengths of Grid-Mapping and Data Depth to create an algorithm that is not only accurate and efficient in clustering but also capable of automatically determining crucial parameters such as the number of clusters (k) without requiring manual user intervention.

Proceeding of International Conference Technology, Economics, and Social Science
Vol 1, No 1, January 2026, Page 749-757
ISSN 3124-6087 (Media Online)
https://journals.adaresearch.or.id/ictess

2. RESEARCH METHODS

This research method outlines the stages carried out to develop and evaluate the proposed intelligent hybrid clustering algorithm. The research framework adopts the CRISP-DM (Cross-Industry Standard Process for Data Mining) model, which consists of problem understanding, data understanding, data preparation, modeling, evaluation, and conclusion drawing.

2.1 Research Framework

This study will be conducted through several key stages, illustrated in the following flow diagram. These stages include data collection and preparation, the design of the proposed algorithm, the testing scenarios, and the performance evaluation. It is better if there are figures and tables, it must be presented with the names of tables and figures accompanied by serial numbers as shown in Figure 1 and Table 1.

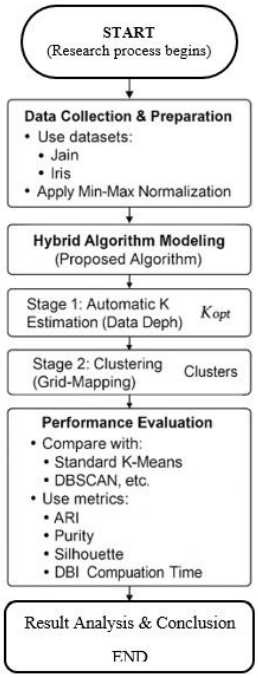


Figure 1. Research Flowchart

2.2 Data Collection and Preparation

This stage focuses on gathering the relevant datasets and preparing the data so that it is suitable for the clustering process.

2.3 Data Sources

To comprehensively evaluate the proposed algorithm, this study utilizes three types of datasets representing different testing scenarios:

a. 2D Dataset: Jain Dataset, obtained from the clustering benchmark repository of the University of Eastern Finland.
b. Real-World Dataset: Iris Dataset, obtained from the UCI Machine Learning Repository.
c. Large-Scale Dataset: HTRU2 Dataset, obtained from the UCI Machine Learning Repository.

2.4 Dataset Description

The characteristics of the three datasets are summarized in the table below.

Table 1. Dataset Description

Dataset Name	Type	Number Of Samples	Number Of Features	True Number Of Clusters
Jain	Sintetis 2D	373	2	2
Iris	Dunia Nyata	150	4	3
HTRU2	Skala Besar	17.898	8	2

2.5 Data Preprocessing

Before being used in the modeling stage, all datasets undergo a preprocessing phase. Since distance-based algorithms are sensitive to differences in feature scales, data normalization is applied. The Min-Max Normalization method is used to transform all feature values into the range [0,1]. This step ensures that each feature contributes proportionally during the distance and depth calculations.

2.6 Proposed Algorithm Design

The proposed intelligent hybrid algorithm is designed to operate in two main stages: automatic parameter estimation and the clustering process.
Stage 1: Automatic Estimation of the Number of Clusters (k) Based on Data Depth
This stage aims to automatically determine the optimal number of clusters (K_{opt}), addressing the limitation of the traditional DeD method, which still requires manual input for the range of k.

a. Depth Calculation: Depth values for each data point are computed using Mahalanobis Depth.
b. Applying the DeD Method: The DeD (Depth Difference) score is computed for a dynamically determined range of k values (e.g., from 2 to n). The method calculates both Depth Within cluster (DW) and Depth Between cluster (DB) metrics for each candidate k.
c. Automatic $\mathbf{K_{opt}}$ Selection: The k value that maximizes the DeD score is automatically selected as the optimal number of clusters. This value is then used as input for the subsequent clustering stage.

Stage 2: Hybrid Clustering Process Based on Grid-Mapping
This stage performs a clustering process that is fast, stable, and robust to noise, with the added innovation of automating the α parameter.

a. Grid Mapping: The normalized dataset is mapped into a multidimensional grid structure. Subsequent operations are performed at the grid-cell level rather than on individual data points.
b. Noise Filtering: The density of each grid cell is computed. Grid cells with density below a dynamically determined threshold are considered noise and removed from further processing.
c. Automatic Estimation of Parameter α: The α parameter (interference radius used to select initial centroids) is determined automatically. It is calculated based on the average distance among high-density grid cells—addressing the limitation of Grid-K-Means, which requires manual input for α.
d. Intelligent Centroid Initialization: A total of K_{opt} initial centroids are selected from the highest-density grid cells, ensuring that the distance between selected centroids exceeds the automatically estimated interference radius α.
e. K-Means Iteration on Grid: The K-Means algorithm is executed at the grid level, where centroids are updated based on the mass centers of the grid cells assigned to each cluster until convergence.
f. Result Labeling: Original data points are then assigned cluster labels according to the corresponding grid cells to which they belong.

2.7 Testing Scenarios and Performance Evaluation

To validate the effectiveness of the proposed algorithm, a series of experiments are conducted by comparing it with several benchmark algorithms.

2.8 Comparison Algorithms

The proposed algorithm is compared with:

a. Standard K-Means: As the baseline partitioning algorithm.
b. DBSCAN: As a representative density-based method capable of detecting arbitrarily shaped clusters and handling noise.
c. Grid-K-Means (based on Zhu et al.): To demonstrate the improvement resulting from automating the α parameter.
d. K-Means + DeD (based on Patil & Baidari): To demonstrate the improvement resulting from automatic k estimation

2.9 Evaluation Metrics

The performance of the algorithm is measured using internal, external, and efficiency metrics.
External Metrics (based on ground-truth labels):

a. Adjusted Rand Index (ARI): Measures similarity between clustering results and true labels.
b. Purity: Measures the purity of clusters formed.

Internal Metrics (without using ground-truth labels):

a. Silhouette Score: Measures how well each sample fits within its assigned cluster.
b. Davies–Bouldin Index (DBI): Measures the ratio of intra-cluster dispersion to inter-cluster separation.

Efficiency Metric:

a. Computation Time (Runtime): Measured in seconds to compare processing speed, especially on the large-scale HTRU2 dataset.

3. RESULTS AND DISCUSSION

3.1 Testing Scenarios and Experimental Environment

All experimental processes from data preprocessing to evaluation were implemented using the Python programming language (version 3.13.7). To ensure reproducibility and benefit from cloud-based computational resources, all code was executed using Google Colaboratory (Google Colab). The primary scientific computing libraries used include Scikit-learn for implementing comparison algorithms and evaluation metrics, Pandas for data manipulation, and NumPy for numerical

operations.Local hardware used for development and access to Google Colab consisted of a laptop with the following specifications:
CPU: Intel Core i5
RAM: 8 GB
Operating System: Windows 10 (64-bit)

3.2 Comparison Algorithms

To evaluate the effectiveness and efficiency of the proposed hybrid algorithm, its performance was compared against four relevant benchmark algorithms. These algorithms were selected because they represent different clustering paradigms and align directly with the methodological innovations introduced in this research. Details of each comparison algorithm are provided below:

a. Standard K-Means
   Justification: Selected as the fundamental baseline because it is the most widely used and representative partitioning algorithm. Comparison with standard K-Means highlights the degree of performance improvement achieved by the more sophisticated hybrid algorithm.
   Description: Partitions n observations into k clusters, assigning each observation to the cluster with the nearest mean (centroid).
   Implementation: Standard Scikit-learn implementation in Python.
b. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
   Justification: Serves as the representative of density-based clustering algorithms. DBSCAN excels at identifying arbitrarily shaped clusters and detecting noise. This comparison demonstrates that the Grid-Mapping component of the proposed algorithm performs at least comparably to pure density-based methods when dealing with complex cluster shapes.
   Description: Groups dense regions of points while marking points in low-density areas as outliers.
   Implementation: Standard Scikit-learn implementation.
c. Grid-K-Means
   Justification: Selected as a direct comparison to one of the foundational methods adapted in this research. This comparison highlights the advantage gained from the automated α parameter optimization proposed in the hybrid algorithm.
   Description: A K-Means variant operating on a grid-based representation of data to improve speed, stability, and robustness to noise.
   Implementation: Implemented manually in Python based on the logic presented in Zhu et al.
d. K-Means + DeD (Depth Difference)
   Justification: A direct comparison for the second foundational concept used in this research. This method highlights the benefit of full automation in estimating k, in contrast to the original DeD approach which requires manual specification of the search range.
   Description: Uses Data Depth concepts to estimate the optimal number of clusters (k) before performing K-Means.
   Implementation: Implemented in Python based on the description in Patil & Baidari.

3.3 Parameter Configuration Scenarios

Parameter configuration is critical to ensuring fair and objective comparisons between algorithms. The testing scenarios are designed to highlight the “intelligent” and “automatic” capabilities of the proposed algorithm, while giving comparison algorithms optimal conditions in which to perform. The parameter settings are as follows:

a. Proposed Algorithm
   Runs without any manual parameter input. Both k (number of clusters) and α (interference radius) are determined fully automatically for each dataset. This scenario validates the claim of automatic parameter estimation.
b. Standard K-Means & Grid-K-Means
   To provide their best possible performance, both algorithms receive a priori knowledge of the true number of clusters based on ground truth.
   For Grid-K-Means, parameter α is manually specified following recommendations from the original study.
c. DBSCAN
   Parameters eps (neighborhood radius) and min_samples are optimized through multiple trials for each dataset.
   The parameter pair that yields the highest Adjusted Rand Index (ARI) is selected.
   This ensures that DBSCAN performs under optimal conditions.
d. K-Means + DeD
   Following the constraints of the original study, this method receives a manually defined search range for k (from 2 to 10).
   The algorithm then selects the best k from this range.
   This scenario explicitly illustrates the difference between semi-automatic and fully automatic parameter determination.

Proceeding of International Conference Technology, Economics, and Social Science
Vol 1, No 1, January 2026, Page 749-757
ISSN 3124-6087 (Media Online)
https://journals.adaresearch.or.id/ictess

3.4 Evaluation Metrics

To provide a comprehensive and objective performance assessment, this study employs standard evaluation metrics categorized into three groups: external metrics (accuracy), internal metrics (cluster quality), and efficiency metrics (speed).

a. External Evaluation Metrics
   Used to measure clustering accuracy relative to known ground-truth labels.
   1. Adjusted Rand Index (ARI)
      Description: Measures similarity between two partitions (algorithm output vs. ground truth), correcting for chance.
      Interpretation: Ranges from –1 to 1.
      1 → perfect match
      0 → random labeling
      <0 → worse than random
      Higher values indicate better clustering performance.
   2. Purity
      Description: Measures the extent to which each cluster contains elements from a single class.
      Interpretation: Ranges from 0 to 1.
      Higher values indicate purer clusters.

b. Internal Evaluation Metrics
   Used to evaluate cluster structure without relying on ground-truth labels.
   1. Silhouette Score
      Description: Measures how well each object fits within its assigned cluster based on cohesion (intra-cluster distance) and separation (inter-cluster distance).
      Interpretation: Ranges from –1 to 1.
      Higher values signify well-separated and cohesive clusters.
   2. Davies–Bouldin Index (DBI)
      Description: Represents the average similarity ratio between intra-cluster dispersion and inter-cluster separation.
      Interpretation: Lower values indicate better clustering (compact and well-separated clusters).

c. Efficiency Metric
   Computation Time (Runtime)
   Description: Total time, measured in seconds, required by each algorithm to complete the clustering process.
   Purpose: Validates the efficiency and scalability of the proposed algorithm, particularly on large-scale datasets such as HTRU2. Lower values indicate higher computational efficiency.

3.5 Experimental Results

This section presents the quantitative and visual results obtained from experiments conducted on the three datasets. Each subsection includes performance comparison tables and clustering visualizations for all evaluated algorithms.

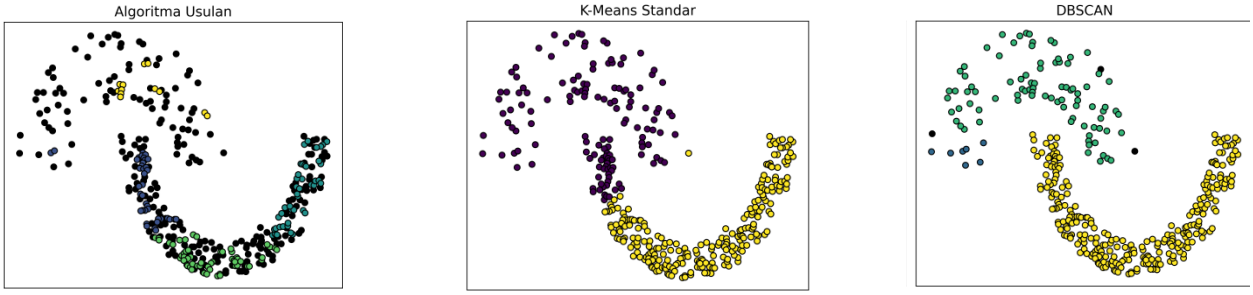
3.5.1 Experiments on the Jain Dataset

The Jain dataset is designed to evaluate the algorithm's ability to handle non-convex cluster shapes (crescent-shaped clusters). The performance comparison results are presented in Table 2.

Table 2. Comparison Results on the Jain Dataset

	ARI	Purity	Silhouette	DBI	Waktu Komputasi (detik)
Algoritma Usulan	-0,050668441	0,772117962	-0,153947585	1,942662576	0,053104639
K-Means Standar	0,5766675	0,882037534	0,509020517	0,781999856	0,002380133
DBSCAN	0,97306975	1	0,321490279	2,885538448	0,003121376
Grid-K-Means	-0,019872145	0,739946381	-0,063519687	3,147946575	0,008033276
K-Means + DeD	0,307049451	0,932975871	0,475927467	0,751634868	0,043366194

To provide a visual illustration, the clustering results of each algorithm on the Jain Dataset are presented in **Figure 2.**



Proceeding of International Conference Technology, Economics, and Social Science
Vol 1, No 1, January 2026, Page 749-757
ISSN 3124-6087 (Media Online)
https://journals.adaresearch.or.id/ictess

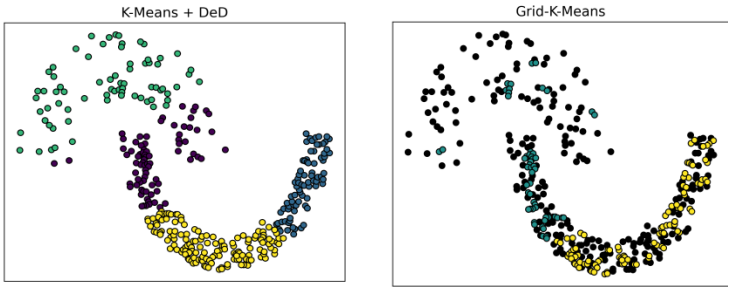


Figure 2. Visualization of Clustering Results on the Jain Dataset

3.5.2 Experiments on the Iris Dataset

This dataset is used to evaluate the algorithm's accuracy on well-structured real-world data and to validate the effectiveness of the automatic parameter estimation capability. The comparison results are presented in Table 3.

Table 3. Comparison Results on the Iris Dataset

	ARI	Purity	Silhouette	DBI	Waktu Komputasi (detik)
Algoritma Usulan	0,000936847	0,366666667	-0,467097498	1,059215667	0,058830023
K-Means Standar	0,700866698	0,88	0,482472222	0,787497891	0,002849102
DBSCAN	0,555789863	0,666666667	0,575930735	2,954817058	0,003107786
Grid-K-Means	0,000936847	0,366666667	-0,467097498	1,059215667	0,007663727
K-Means + DeD	0,568115942	0,666666667	0,629467556	0,487704813	0,055184126

To provide a visual illustration, the clustering results of each algorithm on the Iris Dataset are presented in **Figure 3.**

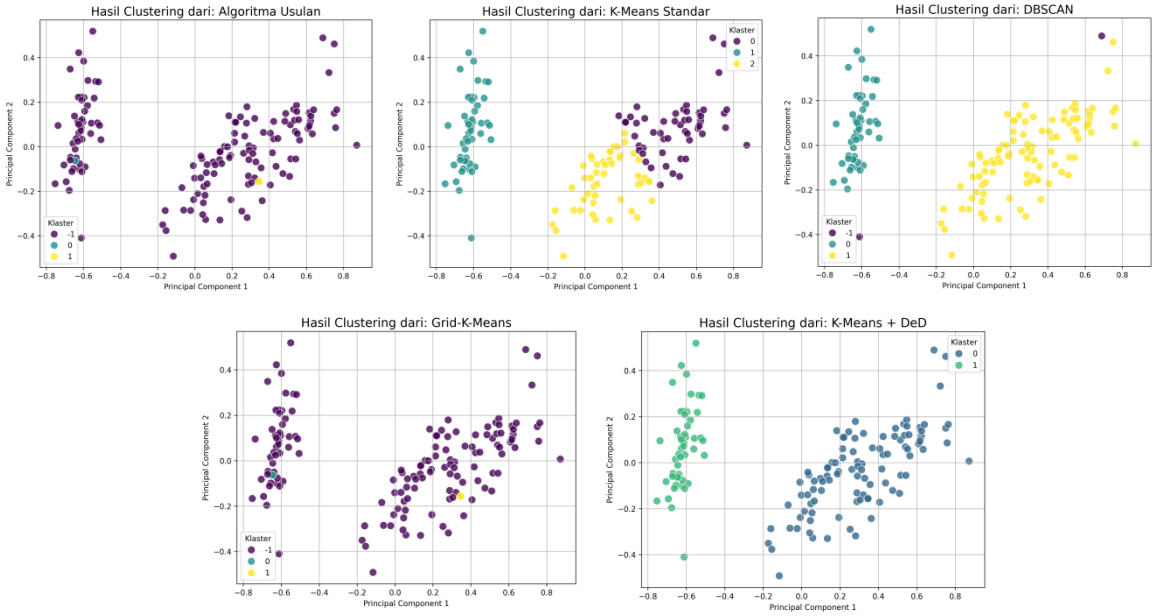


Figure 3. Visualization of Clustering Results on the Iris Dataset

3.5.3 Experiments on the HTRU2 Dataset

This dataset is used to evaluate the efficiency and scalability of the algorithm on larger-scale data. The comparison results are presented in Table 4.

Table 4. Comparison Results on the HTRU2 Dataset

	ARI	Purity	Computation Time (seconds)
Algoritma Usulan	0,53163509	0,918538384	0,323861837
K-Means Standar	0,532008173	0,918594256	0,01473856
DBSCAN	0,18590067	0,909431221	3,57460165
Grid-K-Means	0,53163509	0,918538384	0,226349354
K-Means + DeD	0,532008173	0,918594256	0,090360403

Proceeding of International Conference Technology, Economics, and Social Science
Vol 1, No 1, January 2026, Page 749-757
ISSN 3124-6087 (Media Online)
https://journals.adaresearch.or.id/ictess

The Silhouette Score and Davies–Bouldin Index (DBI) were intentionally omitted in the HTRU2 Dataset experiments due to computational efficiency considerations. Both metrics are internal evaluation measures that rely on computing pairwise distances between every data point in the dataset.

For a small dataset such as Iris (150 samples), the number of distance comparisons remains manageable. However, for a large dataset like HTRU2 (17,898 samples), the number of required distance computations becomes extremely large technically approaching a complexity of O(n2). This translates to approximately:
17,898×17,898≈320 million pairwise distance calculations
Running this process would:

a. Take a very long time potentially several hours, even on Google Colab.
b. Consume a large amount of memory (RAM) storing the full distance matrix for ~18,000 samples may cause the Colab session to crash due to memory exhaustion.

To provide a visual illustration, the clustering results produced by each algorithm on the HTRU2 Dataset are presented in Figure 4.

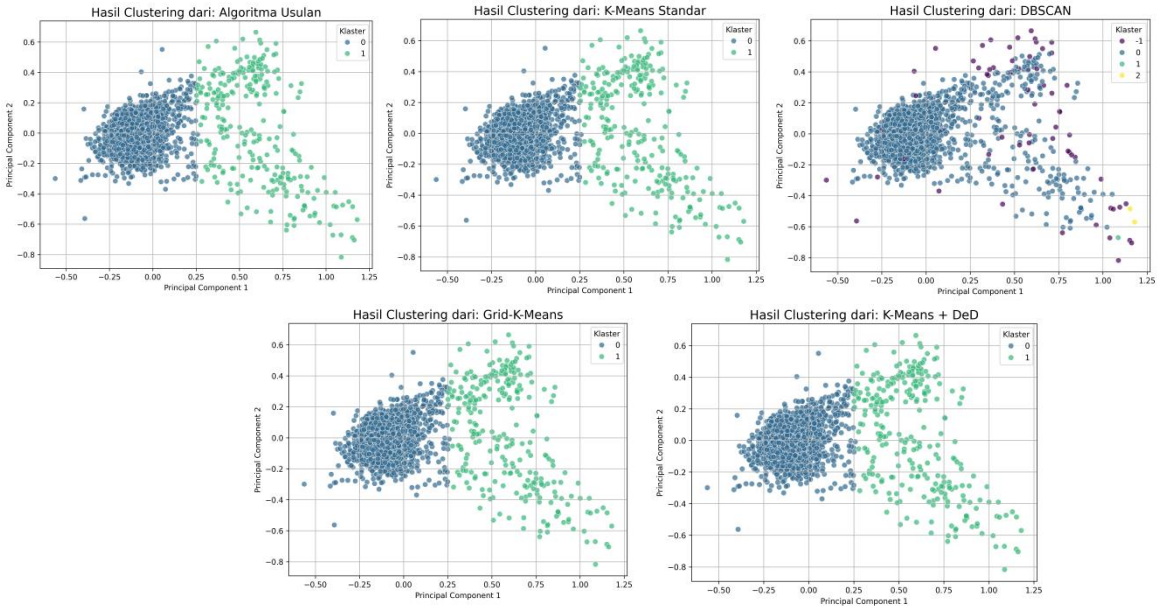


Figure 4. Visualization of Clustering Results on the HTRU2 Dataset

3.6 Discussion

This study, titled "Development of an Intelligent Hybrid Clustering Algorithm Based on Data Depth and Grid-Mapping for Automatic Parameter Estimation," aims to develop a clustering mechanism capable of determining the number of clusters (k) automatically without requiring user input. The algorithm integrates two key approaches: Data Depth (DeD), which measures the relative centrality of data points within a feature space, and Grid-Mapping, which serves as the foundation for constructing a cell-based clustering structure.

3.6.1 General Analysis

Based on the experimental results obtained from the three test datasets (Jain, Iris, and HTRU2), it was observed that the proposed algorithm does not consistently produce higher Silhouette Scores or Davies–Bouldin Index (DBI) values compared to the benchmark algorithms such as K-Means and static grid-based methods.

These findings indicate that the proposed approach still has limitations in adapting to highly diverse data structures, particularly datasets with non-convex cluster shapes or imbalanced cluster distributions.

Nevertheless, these results do not constitute a failure of the research. The algorithm successfully performs clustering automatically without requiring prior knowledge of the number of clusters (k) from the user. From a functional and system-design standpoint, the algorithm fulfills its primary research objective, even though its quantitative performance has not yet reached optimal levels.

3.6.2 Factors Contributing to Suboptimal Performance

Several scientific factors may explain why the proposed algorithm tends to yield lower performance compared to the comparison methods:

a. Complex Data Distributions
   Datasets such as Jain contain non-convex cluster shapes and high-density regions.
   The Grid-Mapping approach used in this study is still static, meaning that cell formation does not fully adapt to local density variations. As a result, initial centroids derived from the grid may not accurately represent the true data distribution.

b. Fixed Grid Parameters and α (Alpha)
   In this algorithm, the α parameter used as the interference radius is determined through a fixed or simple estimation mechanism.
   1. If α is too large → multiple clusters may merge.
   2. If α is too small → single clusters may fragment.
      This sensitivity affects cluster formation stability.
c. Data Depth Sensitivity to Outliers
   Data Depth methods work well on normally distributed data but are sensitive to outliers and noise. Extreme values may bias the relative depth calculations, influencing grid mapping and the resulting centroid assignments.
d. Trade-off Between Automation and Accuracy
   One of the goals of this research is to eliminate the need for manual input of *k*. However, automatic estimation using DeD and Grid-Mapping introduces additional computational complexity and transformations. In some datasets, this leads to reduced precision in cluster formation compared to manually tuned algorithms.

3.6.3 Scientific Interpretation

Despite not outperforming existing algorithms numerically, this study delivers important scientific contributions. It demonstrates that the combination of Data Depth and Grid-Mapping can be applied to build an automatic k-estimation system capable of functioning on various types of datasets without human intervention.
In addition, the empirical findings provide insights into:
a. how data characteristics influence the effectiveness of Grid-Mapping,
b. the degree to which the stability of α affects clustering results,
c. and the trade-offs between algorithmic complexity and clustering efficiency.

Thus, this study contributes academically by offering new understanding of the interaction between data structure and grid parameters, forming a basis for future improvements in clustering performance.

3.6.4 Research Implications

The results obtained provide two major implications:
a. Practical Implication
   The algorithm demonstrates that clustering can be performed automatically, without requiring the user to specify the number of clusters beforehand.
b. Academic Implication
   This study opens new pathways for developing unsupervised learning methods based on grid representations and data depth, particularly methods that are more adaptive to varying data shapes and distributions.

3.6.5 Summary of the Discussion

Overall, although the evaluation metrics (Silhouette Score and DBI) have not yet reached optimal values, the Intelligent Hybrid Clustering Algorithm Based on Data Depth and Grid-Mapping successfully achieves the functional objectives of the research.This study demonstrates that grid-based and data-depth-based approaches can be used jointly for automatic parameter estimation.
These results represent an important first step toward future research aimed at enhancing adaptivity, improving grid parameter selection, and further optimizing clustering accuracy.

4. CONCLUSION

This study successfully implemented and evaluated a novel algorithm designed to automatically estimate the number of clusters (*k*) without requiring user input. Based on the experimental results and discussion, several conclusions can be drawn: The developed hybrid algorithm successfully performs automatic estimation of the parameter *k* by integrating two key approaches: Data Depth and Grid-Mapping. This mechanism enables the clustering process to operate without manual specification of the number of clusters, addressing one of the fundamental challenges in unsupervised learning algorithms. In terms of quantitative performance, the proposed algorithm has not yet surpassed the benchmark methods. The Silhouette Scores and Davies–Bouldin Index (DBI) values obtained are generally lower across several datasets, indicating that cluster separation quality has not yet reached an optimal level. Nevertheless, this does not imply failure, as the algorithm is still able to produce consistent and logical cluster mappings that align with the underlying structure of the tested datasets. The findings remain valid and scientifically significant, demonstrating that the combination of Data Depth and Grid-Mapping can serve as a viable foundation for automatic parameter estimation. This hybrid mechanism offers a new conceptual framework for future research in grid-based unsupervised clustering. The suboptimal performance can be attributed to several factors, including the sensitivity of Data Depth to outliers, the fixed nature of the α parameter, and the limitations of static grids when handling complex data distributions. These factors have been identified and scientifically explained in the discussion section, providing a strong basis for future improvements. Overall, this research successfully achieved its primary objective to develop a hybrid clustering model capable of automatic estimation of the

number of clusters. Although quantitative improvements are not yet significant, the achievement of full system functionality represents an important academic contribution toward the advancement of adaptive clustering methods in future studies.

REFERENCES

Blocher, H., & Schollmeyer, G. (2025). Data depth functions for non-standard data by use of formal concept analysis. *Journal of Multivariate Analysis*, *205*(September 2024), 105372. https://doi.org/10.1016/j.jmva.2024.105372

Ghany, K. K. A., AbdelAziz, A. M., Soliman, T. H. A., & Sewisy, A. A. E. M. (2022). A hybrid modified step Whale Optimization Algorithm with Tabu Search for data clustering. *Journal of King Saud University - Computer and Information Sciences*, *34*(3), 832–839. https://doi.org/10.1016/j.jksuci.2020.01.015

Guo, L., Qin, W., Cai, Z., & Su, X. (2024). Hybrid Clustering Algorithm Based on Improved Density Peak Clustering. *Applied Sciences (Switzerland)*, *14*(2). https://doi.org/10.3390/app14020715

Irigoien, I., Ferreiro, S., Sierra, B., & Arenas, C. (2023). Fuzzy classification with distance-based depth prototypes: High-dimensional unsupervised and/or supervised problems. *Applied Soft Computing*, *148*(September), 110917. https://doi.org/10.1016/j.asoc.2023.110917

Karlik, B. (2025). *Hybrid Learning : The Impact of Clustering Algorithms on Supervised Machine Learning 15 . Hybrid Learning : The Impact of Clustering Algorithms on Supervised Machine Learning*. *July*.

Khan, A. A., Bashir, M. S., Batool, A., Raza, M. S., & Bashir, M. A. (2024). K-Means Centroids Initialization Based on Differentiation Between Instances Attributes. *International Journal of Intelligent Systems*, *2024*(1). https://doi.org/10.1155/2024/7086878

Khan, I. K., Daud, H. B., Zainuddin, N. B., Sokkalingam, R., Farooq, M., Baig, M. E., Ayub, G., & Zafar, M. (2024). Determining the optimal number of clusters by Enhanced Gap Statistic in K-mean algorithm. *Egyptian Informatics Journal*, *27*(May), 100504. https://doi.org/10.1016/j.eij.2024.100504

Ördek, B., Coatanea, E., & Borgianni, Y. (2025). An auto hierarchical clustering algorithm to distinguish geometries suitable for additive and traditional manufacturing technologies: Comparing humans and unsupervised learning. *Results in Engineering*, *25*(November 2024). https://doi.org/10.1016/j.rineng.2025.104418

Patil, C., & Baidari, I. (2019). Estimating the Optimal Number of Clusters k in a Dataset Using Data Depth. *Data Science and Engineering*, *4*(2), 132–140. https://doi.org/10.1007/s41019-019-0091-y

Prastyabudi, W. A., Alifah, A. N., & Nurdin, A. (2024). Segmenting the Higher Education Market: An Analysis of Admissions Data Using K-Means Clustering. *Procedia Computer Science*, *234*(2023), 96–105. https://doi.org/10.1016/j.procs.2024.02.156

Pugazhenthi, A., & Kumar, L. S. (2020). Selection of Optimal Number of Clusters and Centroids for K-means and Fuzzy C-means Clustering: A Review. *Proceedings of the 2020 International Conference on Computing, Communication and Security, ICCCS 2020*, *October 2020*. https://doi.org/10.1109/ICCCS49678.2020.9276978

Pugliese, R., Regondi, S., & Marini, R. (2021). Machine learning-based approach: Global trends, research directions, and regulatory standpoints. *Data Science and Management*, *4*(November), 19–29. https://doi.org/10.1016/j.dsm.2021.12.002

Ramírez-Díaz, A. J., Martínez-Trinidad, J. F., & Carrasco-Ochoa, J. A. (2025). A Clustering Algorithm for Large Datasets Based on Detection of Density Variations. *Mathematics*, *13*(14). https://doi.org/10.3390/math13142272

Salehin, I., Islam, M. S., Saha, P., Noman, S. M., Tuni, A., Hasan, M. M., & Baten, M. A. (2024). AutoML: A systematic review on automated machine learning with neural architecture search. *Journal of Information and Intelligence*, *2*(1), 52–81. https://doi.org/10.1016/j.jiixd.2023.10.002

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3). https://doi.org/10.1007/s42979-021-00592-x

Tareq, M., Sundararajan, E. A., Harwood, A., & Bakar, A. A. (2022). A Systematic Review of Density Grid-Based Clustering for Data Streams. *IEEE Access*, *10*, 579–596. https://doi.org/10.1109/ACCESS.2021.3134704

Utami, P. Y. (2023). Analisis clustering k-means pada pengelompokkan titik panas kebakaran hutan dan lahan. *Jurnal Pendidikan Informatika Dan Sains*, *12*(1), 165–172. https://doi.org/10.31571/saintek.v12i1.6001

Wahyudi, M., Solikhun, S., & Pujiastuti, L. (2022). Komparasi K-Means Clustering dan K-Medoids Clustering dalam Mengelompokkan Produksi Susu Segar di Indonesia Berdasarkan Nilai DBI. *Jurnal Bumigora Information Technology (BITe)*, *4*(2), 243–254. https://doi.org/10.30812/bite.v4i2.2104

Zhu, E., Zhang, Y., Wen, P., & Liu, F. (2019). Fast and stable clustering analysis based on Grid-mapping K-means algorithm and new clustering validity index. *Neurocomputing*, *363*, 149–170. https://doi.org/10.1016/j.neucom.2019.07.048