IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Data Clustering: Prospects & Challenges

[1]Md Ashraful Alam, [2]Hammed Esa, [3]Kazi Nazrul Islam, [4]Anwar Hossain

[1]Graduate Research Assistant, [2-4]Graduate Student

[1]Department of Electrical Engineering

[1]Tuskegee University, Tuskegee, USA
[2]International American University, USA
[3]Westcliff University, California, USA
[4]California State University San Bernardino, California, USA

Abstract: Data clustering primarily serves as a solution for tackling unsupervised learning challenges and represents a fundamental tool applied across various domains-including data mining, pattern recognition, and artificial intelligence. The main objective of data clustering is to group similar objects and allocate them to different categories. Different clustering techniques are developed and implemented to categorize data objects. Partitioning and hierarchical techniques [1] are the two classified categories of data clustering. Additionally, alternative methods such as grid-based, density-based, and fuzzy C-mean clustering approaches are also available. This paper's objective is to provide a comprehensive overview of data clustering. It covers the historical context of this technique, furnishes a precise definition, and thoroughly explores various types of clustering methods while critically assessing their respective strengths and weaknesses. Furthermore, the paper delves into the practical applications of data clustering and highlights recently developed algorithms, facilitating a meaningful comparison among different clustering approaches. In summary, this paper offers a concise yet thorough review of the entire spectrum of whole data clustering method.

Keywords— Data clustering, hierarchical clustering, partitioning technique, k-mean.

I. Introduction

In recent years, an enormous amount of data has been generated, with nearly 90% of this data originating within the past two years. Data from 2020 indicates that individuals were producing data at a rate of 1.7 megabytes per second, equating to an average daily total of 2.5 quintillion bytes. Projections for 2025 suggest that an astonishing 463 exabytes of data will be generated each day. In the last three years 2020, 2021, 2022 shows that the world produced 64.2 zettabytes, 79 zettabytes, 97 zettabytes data respectively. Projected data generated in 2023 and 2024 are 120 zettabytes and 147 zettabytes respectively over the world. Managing such colossal data volumes is an enormous task. This data starts in its raw and unstructured form, and to derive valuable information from it, processing is imperative; otherwise, the data remains essentially worthless [2]. However, manually processing this data is practically impossible due to its vastness and complexity. Additionally, programming it directly is infeasible due to the multiple dimensions involved. This is where machine learning, deep learning [3] enters the scene. Deep learning has emerged as a powerful tool in the field of data clustering, offering several advantages over traditional clustering methods. Deep learning-based clustering techniques leverage artificial neural networks [4] with multiple hidden layers to automatically discover complex patterns and representations in data. In machine learning [5], machines are trained using data. There are two primary learning techniques: supervised learning, where data is labeled and static, enabling machines to learn from this labeled data, and unsupervised learning, which deals with unlabeled and dynamic data, where machines are trained to make decisions based on this unlabeled data. However, annotating large datasets can be prohibitively expensive and challenging. For example, achieving a state-of-the-art transcription level in a speech recognition system would necessitate thousands of hours of costly manual effort. Moreover, in many cases, predefined classifications cannot be readily established. In data mining, where vast and diverse datasets are collected, defining clear classes can be exceptionally difficult. Hence, unsupervised learning becomes essential for such scenarios [6]. To extract knowledge from unsupervised data, data clustering emerges as one of the most prevalent and indispensable techniques.

Furthermore, recent years have witnessed a deluge of data that needs to be collected automatically, including weather data, worldwide web data, and dynamic social media data. Decisions are often made based on the analysis of this data. Data clustering is a prominent technique for capturing dynamic data, facilitating their analysis, and aiding in informed decision-making.

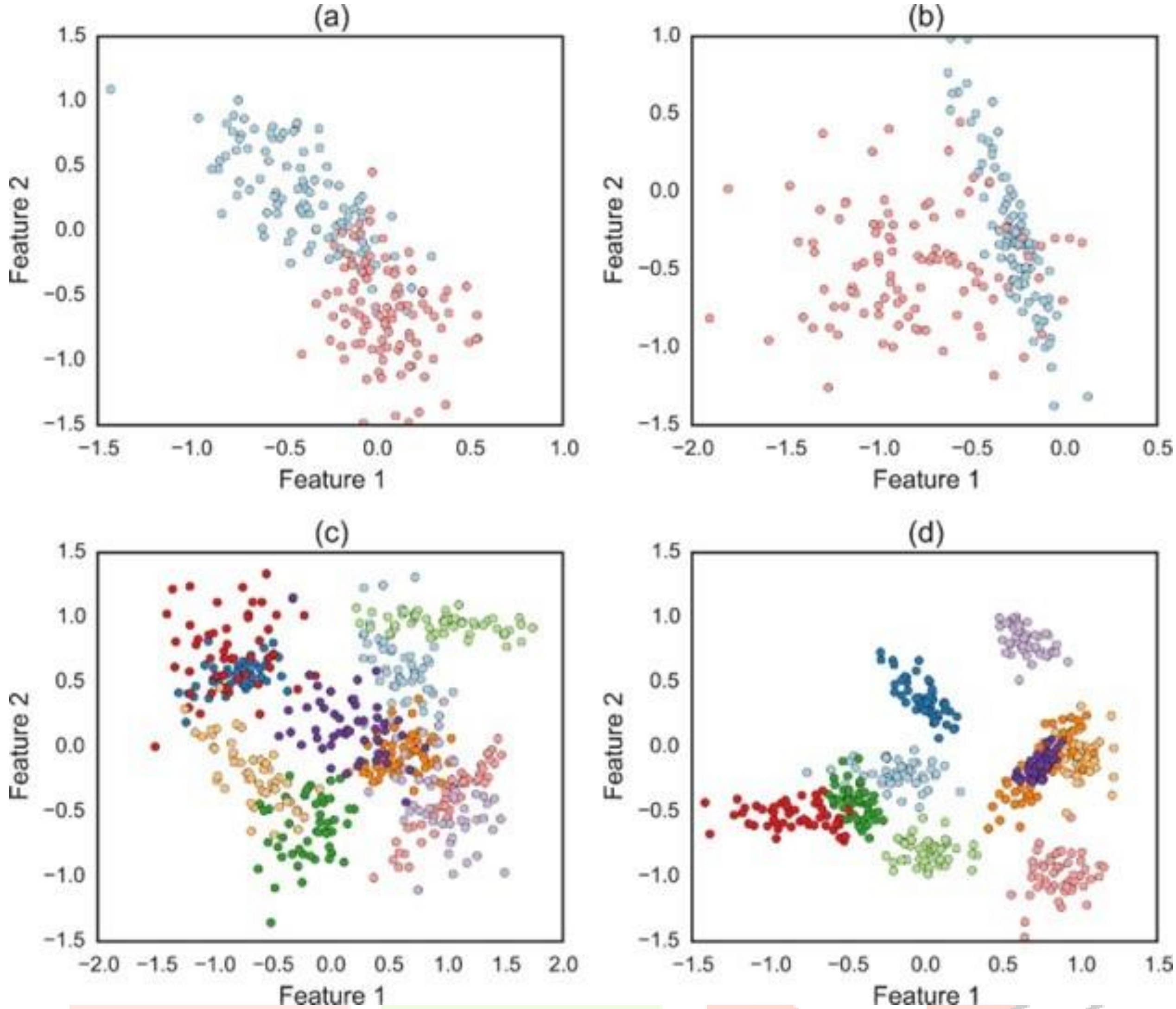

Fig-1: Data Clustering techniques

Data mining involves extracting valuable insights from large datasets, and one of its key features is the ability to handle extensive data sets. However, previously available algorithms were not equipped to efficiently process such vast datasets. Data clustering stands as a fundamental tool to manage and analyze this data [7]. Data clustering categorizes similar entities based on clusters, with the quality of clustering relying on its ability to identify patterns. In a successful cluster, the intra-cluster distance should be minimized, while the inter-cluster distance should be maximized. Various methods are employed in cluster analysis to address different constraints such as result interpretation, outlier detection, handling dynamic data, and evaluating results. These methods aim to mitigate the limitations inherent in cluster analysis. Depending on specific requirements, users select the most suitable clustering method, recognizing that no single method is flawless, as each has its own set of advantages and limitations [8].

Different clustering methods are employed based on the specific requirements of data clustering. Some of the methods rely on principles of distance and similarity as their foundational algorithms. For quantitative data, common distance and similarity functions are used. Following the determination of distance and similarity functions, an evaluation indicator is employed, ultimately making the algorithm suitable for analysis. Clustering is then categorized into various groups based on the specific analysis requirements.

II. CLASIFICATION OF DALA CLUSTERING METHOD

2.1: Hierarchical Clustering Method

Hierarchical clustering is a cluster analysis technique that groups similar objects into clusters, with each cluster being distinct from the others. There are two main types of hierarchical clustering: Agglomerative Hierarchical Clustering (AGNES) and Divisive Hierarchical Clustering (DIANA). Hierarchical clustering is particularly effective for identifying small clusters, while divisive clustering is more suitable for identifying larger clusters.

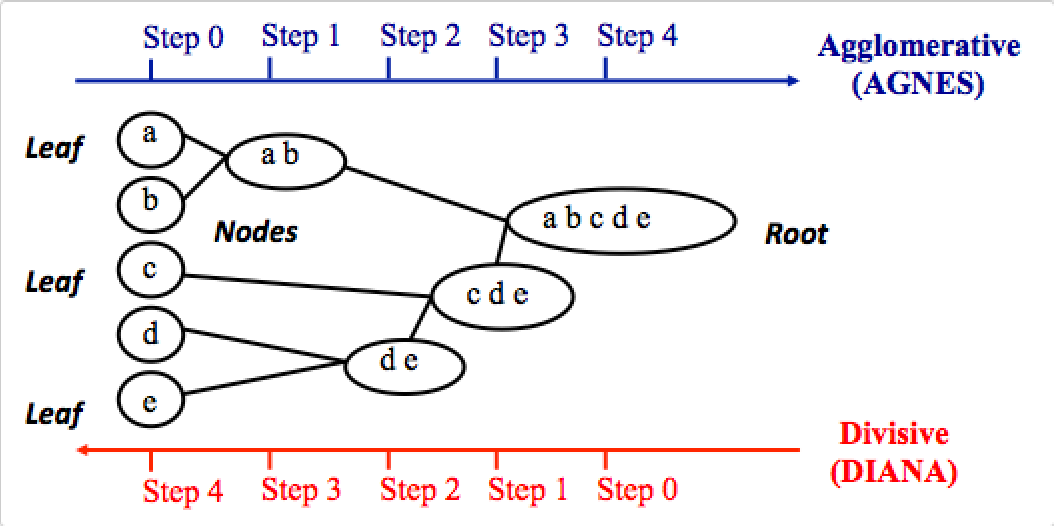

Fig-2: Comparison of agglomerative and divisive hierarchal clustering

In agglomerative hierarchical clustering, the process begins with each object being treated as an individual cluster (leaf), and then the most similar clusters are successively merged to form larger clusters (nodes). This process continues until all objects are members of a single large cluster (root).

Conversely, divisive clustering takes a top-down approach, starting with one large cluster (root) that is successively divided into smaller clusters. The process continues until each object is in its own individual cluster (leaf).

In contemporary applications, advanced agglomerative hierarchical clustering algorithms [9] like CURE (Clustering Using Representatives), BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies), ROCK (Robust Clustering using links), CHEMELEON Algorithm, and various linkage algorithms have been employed to enhance clustering performance and accuracy.

2.1.1: Algorithmic steps for agglomerative hierarchical Clustering

Agglomerative hierarchical clustering starts with each data point as its own cluster and gradually merges clusters until a single cluster containing all data points is formed. The steps involved are as follows:

1) Initialization: Begin by considering each data point as an individual cluster. You have as many clusters as data points initially.
2) Distance Calculation: Compute the pairwise distances between all clusters or data points in the dataset. Common distance metrics include Euclidean distance, Manhattan distance, or other appropriate measures depending on the data type.
3) Merge Closest: Identify the two clusters (or data points) with the smallest distance between them. Merge these clusters into a single cluster. This step reduces the total number of clusters by one.
4) Update Distance Matrix: Recalculate the pairwise distances between the newly formed cluster and all remaining clusters (including individual data points).
5) Repeat: Repeat steps 3 and 4 iteratively until only one cluster, containing all data points, remains.
6) Hierarchical Representation: Throughout this process, you can build a hierarchical tree-like structure called a dendrogram, which visually represents the merging process. It helps in selecting the number of clusters at a later stage by cutting the dendrogram at a suitable height.

2.1.2: Algorithmic steps for divisive hierarchical clustering

Divisive hierarchical clustering starts with a single cluster containing all data points and recursively splits it into smaller clusters. The steps are as follows:

1) Initialization: Begin with a single cluster that contains all data points.
2) Distance Calculation: Compute the pairwise distances between all data points within the current cluster.
3) Split into Two: Identify a point within the current cluster where splitting would be most appropriate. Common methods include k-means or other clustering algorithms. Split the cluster into two smaller clusters based on this criterion.
4) Update Distance Matrix: Recalculate the pairwise distances between the newly formed clusters and all remaining clusters (if any).
5) Repeat: Continue to split clusters recursively until each data point is in its individual cluster or until another stopping criterion is met.

6) Hierarchical Representation: Similar to agglomerative clustering, you can build a dendrogram to visualize the hierarchical structure.

2.1.3: Advantage and Disadvantage of Hierarchical Clustering

The primary advantage of the hierarchical clustering method is its simplicity in implementation and its independence from prior information. Consequently, it often yields superior results compared to other clustering methods. Additionally, it offers more comprehensive insights than the unstructured collection of flat clusters produced by k-means clustering.
However, there are drawbacks to hierarchical clustering. It is an irreversible process, meaning you cannot backtrack to a previous stage. For extensive datasets, hierarchical clustering can be inefficient due to the time it consumes. Furthermore, determining the correct number of clusters from the dendrogram can occasionally pose a challenge.

2.2: Partitioning Clustering Method

The partitioning clustering approach involves creating M clusters, with each item assigned to its distinct cluster. Through observation, various sets of clusters are selected and analyzed to achieve improved outcomes. To use this algorithm, the analyst must specify the desired number of clusters. In this context, we will explore various commonly employed partitioning-based algorithmic methods.

2.2.1: K-Means clustering

K-Means clustering is a widely used partitioning technique in unsupervised machine learning, employed to group data points into distinct clusters. The primary objective is to partition the data into K clusters, where K is a user-specified parameter. Here's a detailed explanation of the K-Means clustering algorithm:

1) Initialization: The algorithm commences by randomly selecting K initial cluster centroids from the dataset. These centroids serve as the starting points for the clusters.
2) Assignment: Each data point is then assigned to the nearest centroid based on a chosen distance metric, typically the Euclidean distance. This step groups data points into K clusters.
3) Update Centroids: After the assignment phase, the algorithm calculates new centroids for each cluster by computing the mean (average) of all data points within that cluster. These newly calculated centroids represent the center of each cluster.
4) Iterative Process: Steps 2 and 3 are repeated iteratively until convergence. Data points are reassigned to clusters based on the nearest centroids, and centroids are updated accordingly. The process continues until the centroids no longer change significantly, or until a predefined stopping criterion, such as a maximum number of iterations, is met.
5) Final Clustering: Once the algorithm converges, it produces the final clustering, where each data point belongs to one of the K clusters.

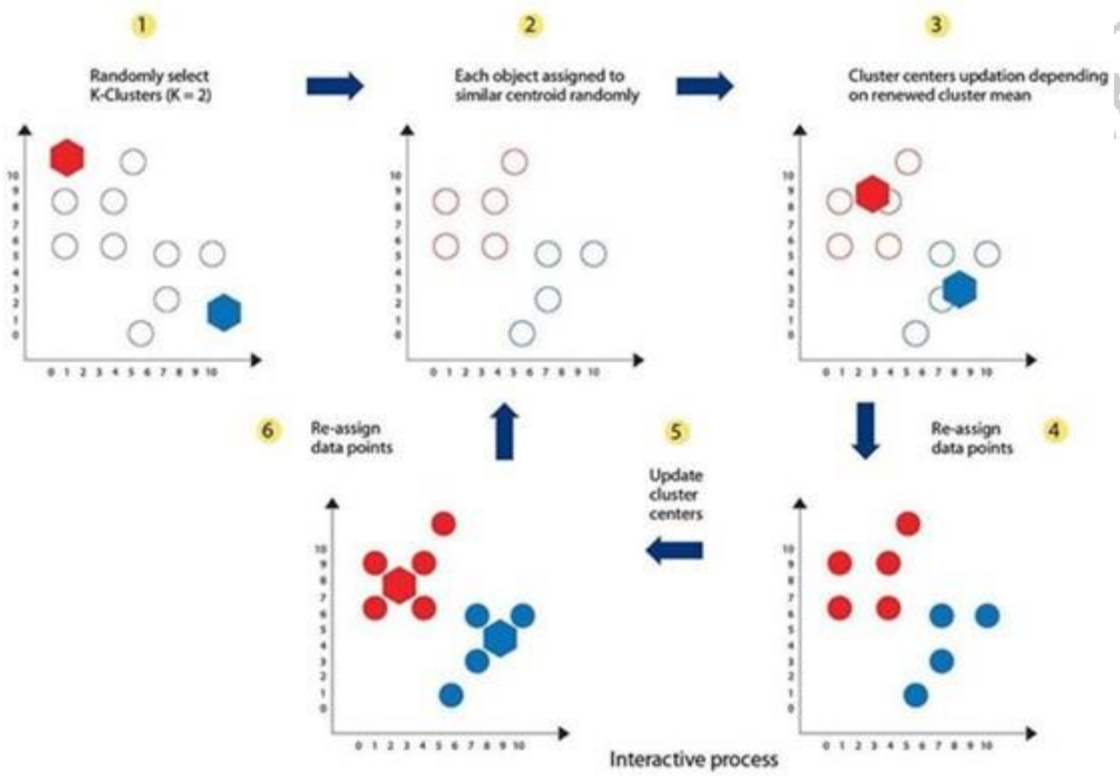

Fig-3: K-Mean method

2.2.2: K-Medoids clustering:

K-Medoids clustering is an extension of the K-Means clustering method, specifically designed to overcome some of its limitations. In K-Medoids clustering, instead of using the mean (average) data point as the centroid for each cluster, it employs an actual data point from the dataset as the representative point, which is known as the "medoid." This makes K-Medoids more robust to outliers and noisy data compared to K-Means.

The algorithm works as follows:

1) Initialization: Like K-Means, K-Medoids begins by initializing K medoids, typically by selecting K data points randomly as the initial medoids.
2) Assignment: In this step, each data point is assigned to the nearest medoid, typically based on a distance metric, often using the Manhattan or Euclidean distance.
3) Update Medoids: After all data points have been assigned to medoids, the algorithm checks if swapping any data point with a non-medoid point as the medoid would decrease the overall cost (commonly measured by the sum of distances within clusters). If such a swap is beneficial, the medoid is updated to the new data point.
4) Repeat Assignment and Update: Steps 2 and 3 are iteratively repeated until convergence. The algorithm continues to update assignments and medoids until there is no further improvement in the clustering or until it reaches a predefined stopping criterion (e.g., a maximum number of iterations).
5) Final Clustering: Once the algorithm converges, it produces the final clustering, with each data point assigned to one of the K clusters, where each cluster is represented by its medoid.

2.2.3: Fuzzy K-Means Clustering:

Fuzzy K-Means (FKM) clustering is an extension of the traditional K-Means clustering algorithm that allows data points to belong to multiple clusters with varying degrees of membership. While traditional K-Means assigns each data point to exactly one cluster, Fuzzy K-Means [22] introduces a soft assignment approach, where data points have membership values indicating their likelihood of belonging to each cluster.

Here are steps in Fuzzy K-Means clustering-
1) Initialization: Start by initializing cluster centroids and membership degrees. The centroids can be initialized randomly or through some other method.
2) Membership Degree Calculation: Calculate the membership degrees for each data point and each cluster using a suitable formula, often based on the distances between data points and cluster centroids. A common formula is based on the Euclidean distance.
3) Update Cluster Centroids: Update the cluster centroids using the membership degrees. Each centroid is recalculated as a weighted average of the data points, where the weights are the membership degrees.
4) Iterative Process: Repeat the membership degree calculation and centroid update steps iteratively until a stopping criterion is met. This can be a maximum number of iterations or when membership degrees and centroids no longer change significantly.
5) Cluster Assignment: Once the algorithm converges, you can assign each data point to one or more clusters based on their membership degrees. Data points are often assigned to clusters with the highest membership degrees.

2.2.4: Advantage and Disadvantage of Partitioning Clustering Method:

In summary, partitioning clustering methods have their strengths in terms of simplicity, scalability, and efficiency, but they also have limitations related to sensitivity to initialization, assumptions about cluster shapes, and the need to specify the number of clusters. Choosing the right clustering method depends on the specific characteristics of the data and the goals of the analysis.

2.3: Density-based Clustering Method

Density-based clustering is a category of clustering algorithms that identify clusters based on the density of data points in the feature space. Unlike partitioning methods like K-Means, density-based methods do not assume that clusters are spherical or equally sized, making them suitable for datasets with irregularly shaped clusters or varying cluster densities. One of the most well-known density-based clustering algorithms is DBSCAN [20] (Density-Based Spatial Clustering of Applications with Noise).

Here are Algorithm Steps (DBSCAN):

1) Parameter Selection: Specify two parameters: the minimum number of data points (MinPts) required to form a dense region and the maximum distance (ε or epsilon) that defines the neighborhood around each data point.
2) Core Point Identification: For each data point, calculate the number of data points within the ε distance. If this count is greater than or equal to MinPts, mark the data point as a core point.
3) Cluster Formation: Start with an arbitrary unvisited core point. Expand the cluster by adding all reachable data points within ε distance to the cluster. Repeat this process for all core points and their reachable neighbors. Each connected component forms a cluster.
4) Border Point Assignment: Assign border points to the cluster of their corresponding core point if they fall within ε distance.
5) Noise Point Identification: Any unvisited data points that are not part of any cluster are considered noise points.

2.3.1: Advantage and Disadvantage of Density-based Clustering Method

Density-based clustering methods like DBSCAN are particularly useful for spatial data analysis, anomaly detection, and applications where clusters have irregular shapes and varying densities. However, proper parameter tuning is essential to obtain meaningful results.

III. APPLICATION OF DATA CLUSTERING

Data clustering finds a wide range of applications across diverse fields due to its ability to reveal hidden patterns and group similar data points, making it a powerful tool for data analysis and decision-making. In the field of marketing, businesses utilize clustering techniques to perform customer segmentation. By categorizing customers based on their buying behaviors, preferences, and demographics, companies can tailor their marketing strategies, advertising campaigns, and product recommendations to specific customer segments. This enhances customer engagement and increases the effectiveness of marketing efforts, ultimately leading to improved customer satisfaction and higher revenues.

In healthcare and bioinformatics [18], data clustering plays a pivotal role in genomics and medical diagnosis. It is used to group genes with similar expression profiles, aiding researchers in identifying potential biomarkers for diseases and understanding complex genetic interactions [9]. The systems can be incorporated to develop biosensors based on carbon-based nanomaterials [10, 11, 12]. In educational settings, data clustering can be employed in active learning strategies to group students based on learning styles or performance, enabling educators to tailor instruction for more effective engagement and understanding [13]. Additionally, in clinical settings, clustering helps categorize patients with similar medical histories or symptoms, facilitating disease diagnosis and personalized treatment plans. Furthermore, in the realm of finance, data clustering is essential for fraud detection. By clustering financial transactions based on transaction patterns, anomalies, and suspicious activities, financial institutions can identify potential fraudulent transactions, protect their clients from fraud, and enhance overall security in the financial sector.

Also, Data clustering plays a crucial role in the modern power system management [3], offering a robust solution for organizing and analyzing the vast amounts of data generated within electrical grids. In power systems, data clustering is employed to group similar assets, such as generators or loads, based on their operational characteristics and performance data. By identifying these clusters, utilities and grid operators can gain valuable insights into the behavior of different components and their interactions. This information is instrumental in load forecasting, fault detection, and predictive maintenance. Moreover, clustering techniques are used to categorize renewable energy [19] sources like wind and solar farms, facilitating their integration into the grid. In the context of smart grids, clustering aids in the identification of demand patterns, allowing for more efficient energy distribution and demand response strategies. Overall, data clustering empowers power system professionals to make informed decisions, enhance grid resilience, and optimize the allocation of resources, ultimately contributing to a more reliable and sustainable electrical infrastructure. Also, clustering techniques have been used to collect relevant data such as voltage, current, temperature, and capacity measurements to improve performance of renewable energy (storage power, lead acid battery). Clustering algorithms like K-means, hierarchical clustering, or DBSCAN to group batteries data with similar characteristics. Once clustered, you can perform specific analyses and develop strategies tailored to each cluster's needs, ultimately improving the reliability and efficiency of lead-acid battery systems [17].

Another significant application of data clustering is in natural language processing and text analysis. Clustering techniques are employed to group similar documents, articles, or social media posts enabling content recommendation systems to suggest relevant articles or posts to users. This enhances user experience on news websites, social media platforms, and content-driven applications. Moreover, in the field of geospatial analysis, cybersecurity [21] clustering assists in identifying spatial patterns and hotspots in various domains, such as disease outbreak prediction, traffic management, and resource allocation in disaster response.

IV. CONCLUSION ANF FUTURE RESEARCH

While significant advancements have been made in data clustering, there is currently no standardized method due to the diverse nature of data. As data types and analysis requirements continue to evolve, researchers are actively working on developing a variety of clustering techniques. Several innovative methods have already been introduced, including:

Black Hole Algorithm: This algorithm draws inspiration from the natural concept of black holes [14]. It operates on the principle of a black hole consuming its closest star. Similarly, in this algorithm, the best-fit data point is treated as a black hole, and the data point closest to it is regarded as a star. The star is then absorbed by the black hole, and this process continues, guiding the clustering process. Notably, this algorithm has shown superior performance compared to traditional clustering methods.

Stratified Sampling-Based Clustering Algorithm: In this approach, a specific number of representative samples are selected from various strata within the dataset. Fuzzy c-means is utilized to select samples from different clusters, and out-of-cluster samples are assigned to their nearest clusters. This technique has proven effective [15], particularly when dealing with complex datasets.

Gravitational Search Algorithm (GSA-LA): GSA-LA combines gravitational search techniques [16] with local search methods to enhance the quality of clustering results. It begins with an initial population and applies the k-means algorithm. Subsequently, the Gravitational Emulation Local Search (GELS) algorithm is employed to refine the clusters. Although it may have higher computational requirements, it ultimately yields more accurate clustering.

K-Mean Clustering with Genetic Algorithm: Recognizing the limitations of traditional K-Means clustering, this approach combines K-Means with genetic algorithms. Genetic algorithms are employed to optimize the initial placement of cluster centroids, resulting in significant improvements in clustering quality.

These innovative approaches exemplify ongoing efforts to address evolving data clustering needs and challenges. As data continues to evolve, along with shifting user expectations and requirements, clustering methods are expected to continuously adapt to meet these dynamic demands. Researchers and practitioners are dedicated to further refining and developing clustering techniques to extract valuable insights from the ever-changing and diverse landscape of data.

V. REFERENCE

[1] C. H. Q. Ding, and, and and H. D. Simon, “A min-max cut algorithm for graph partitioning and data clustering,” in Proceedings 2001 IEEE International Conference on Data Mining, 2001, pp. 107–114.

[2] “Data vs Information - Difference and Comparison | Diffen.” [Online]. Available: https://www.diffen.com/difference/Data_vs_Information. [Accessed: 04-May-2019]

[3] S. U. Khan, M. Mynuddin, D. M. A. Ahad, M. I. Hossain, M. J. Islam and M. F. Kabir, "A Comparative Analysis of Deep Learning Models for Power Quality Disturbance Classification," 2023 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2023, pp. 0317-0323, doi: 10.1109/AIIoT58121.2023.10174320

[4] Md Helal Hossen, (2023). 3D Human Pose Estimation Via Deep Learning Methods. North American Academic Research, 6(2), 72-82. doi: https://doi.org/10.5281/zenodo.7706409.

[5] Md Suhel Rana, Md Humayun Kabir, Abdus Sobur (2023). Comparison of the Error Rates of MNIST DatasetsUsing Different Type of Machine Learning Model. North American Academic Research, 6(5), 173-181. doi:https://doi.org/10.5281/zenodo.8010602

[6] S. Mishra, “Unsupervised Learning and Data Clustering,” Towards Data Science, 19-May-2017. [Online]. Available: https://towardsdatascience.com/unsupervised-learning-and-data-clustering-eeecb78b422a. [Accessed: 04-May-2019].

[7] M. Z. Rodriguez et al., “Clustering algorithms: A comparative approach,” PLOS ONE, vol. 14, no. 1, p. e0210236, Jan. 2019

[8] S. Bano and N. Khan, “A Survey of Data Clustering Methods,” Int. J. Adv. Sci. Technol., vol. 113, Apr. 2018.

[9] D. Xu and Y. Tian, “A Comprehensive Survey of Clustering Algorithms,” Ann. Data Sci., vol. 2, no. 2, pp. 165– 193, Jun. 2015

[10] Sultana, Nigar, Hannah Dewey, and Januka Budhathoki-Uprety. "(Second Place Poster Award) Development of Optical Nanosensors for pH Measurements in Model Biofluids." Electrochemical Society Meeting Abstracts 243. No. 9. The Electrochemical Society, Inc., 2023.

[11] Dewey, Hannah M., Jaron Jones, Sydney Lucas, Shelby Hall, Nigar Sultana, Sofia Mariapaz Abello, and Januka Budhathoki-Uprety. "Carbon Nanotubes for Optical Detection of Quaternary Ammonium Compounds in Complex Media." ACS Applied Nano Materials (2023).

[12] Sultana, Nigar, Hannah Dewey, and Januka Budhathoki-Uprety. "(Second Place Poster Award) Development of Optical Nanosensors for pH Measurements in Model Biofluids." Electrochemical Society Meeting Abstracts 243. No. 9. The Electrochemical Society, Inc., 2023.

[13] Michaelson, Dawn, Nigar Sultana, and Karla P. Teel. "Active learning spaces: Student perceptions of engagement, space, and instructor involvement in an apparel production and merchandising course." International Textile and Apparel Association Annual Conference Proceedings. Vol. 75. No. 1. Iowa State University Digital Press, 2018.

[14] A. Hatamlou, “Black hole: A new heuristic optimization approach for data clustering,” Inf. Sci., vol. 222, pp. 175–184, Feb. 2013.

[15] X. Zhao, J. Liang, and C. Dang, “A stratified sampling based clustering algorithm for large-scale data,” Knowl.- Based Syst., vol. 163, pp. 416–428, Jan. 2019.

[16] T. Mohammadpour, A. M. Bidgoli, R. Enayatifar, and H. H. S. Javadi, “Efficient clustering in collaborative filtering recommender system: Hybrid method based on genetic algorithm and gravitational emulation local search algorithm,” Genomics, Jan. 2019.

[17] A. Fattah, M. A. Kashem and M. J. Islam, "Time Reduction of Deep Cycle Lead Acid Battery Negative Plate Curing Process by Changing Curing Parameters," 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, 2019, pp. 1-6, doi: 10.1109/EICT48899.2019.9068853.

[18] Sharmin Sultana, Md Helal Hossen, Md Jahidul Islam, Jubayer Ahmed, Afsana Mou (2023). Detection, reduction, and filtration of cancer cells through a new DNA polymerization sequence approach. North American Academic Research, 6(1), 230-239. doi: https://doi.org/10.5281/zenodo.7641213.

[19] Md Jahidul Islam, Tanvir Hassan Mojumder, Mushfiqur Rahman, S.M. Nasif Shams (2023). Performance Evaluation of Magnifying Glass and Reflector Based Solar Cooker. North American Academic Research, 6(8), 32-41. doi:https://doi.org/10.5281/zenodo.832227.

[20] Ruiz, C., Spiliopoulou, M., Menasalvas, E. (2007). C-DBSCAN: Density-Based Clustering with Constraints. In: An, A., Stefanowski, J., Ramanna, S., Butz, C.J., Pedrycz, W., Wang, G. (eds) Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. RSFDGrC 2007. Lecture Notes in Computer Science(), vol 4482. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72530-5_25.

[21] Nazrul Islam, Kazi and Sobur, Abdus and Kabir, Md Humayun, The Right to Life of Children and Cyberbullying Dominates Human Rights: Society Impacts (August 8, 2023). Available at SSRN: https://ssrn.com/abstract=4537139 or http://dx.doi.org/10.2139/ssrn.4537139

[22] V. K. Dehariya, S. K. Shrivastava and R. C. Jain, "Clustering of Image Data Set Using K-Means and Fuzzy K-Means Algorithms," 2010 International Conference on Computational Intelligence and Communication Networks, Bhopal, India, 2010, pp. 386-391, doi: 10.1109/CICN.2010.80.