



Integrating Al-Khawarizmi's Principles in K-Means Algorithm: A Case Study of Household Electricity Access Data in Indonesia

Arinal Haq Fauziah

Universitas Islam Negeri Madura, Pamekasan, Indonesia
arinaafauzi@gmail.com

Abstract. This study analyzes the integration of Al-Khawarizmi's algorithmic principles into the K-Means clustering algorithm through a case study of household electricity access in Indonesia (2024). Employing a quantitative approach and algorithmic analysis, secondary data from BPS were classified using K-Means to group regions based on their levels of electricity accessibility. The findings reveal that key principles inherited from Al-Khawarizmi—such as systematic reasoning, structural simplification, applied rationality, and a quantitative-logical framework—contribute significantly to shaping the procedural logic of K-Means, enhancing the clarity and efficiency of data classification. This integration not only improves segmentation quality and result interpretability but also introduces a transhistorical perspective that bridges classical Islamic scientific heritage with contemporary data analysis practices. The study further contributes to ethical and epistemological discourse by emphasizing the incorporation of distributive justice ('*adl*) into algorithmic design. These insights advance the development of a cross-civilizational, value-based computing paradigm and open avenues for future research into algorithms that are both technically robust and socially equitable.

Keywords: Al-Khawarizmi, K-Means algorithm, data clustering.

1 Introduction

Science is an essential foundation in supporting the sustainability of human life, which continues to evolve. The various branches of science known today cannot be separated from the role of scientists who have formulated and systematically organised theories. In the historical narrative of science that has developed over the past few centuries, the dominance of Western scientists, particularly from Europe, is often highlighted as the primary pioneers in fields such as mathematics, astronomy, physics, and medicine. Modern literature often highlights figures such as Isaac Newton, Galileo Galilei, and René Descartes as symbols of human intellectual progress, thereby creating the perception that the entire foundation of modern science stems from Western scientific traditions [1].

However, this perception overlooks the significant contributions of classical Muslim scientists such as Al-Khawarizmi, who, in various scientific historical studies, is more deserving of being called the father of algebra. In his work *Al-Kitāb al-Mukhtaṣar fī*

Ḥisāb al-Jabr wal-Muqābalah, Al-Khawarizmi not only systematically structured algebra but also applied it in real-world contexts such as inheritance calculations, trade, and measurement [2]. With a rational, applied, and systematic approach, he laid the foundation for algorithmic models that remain relevant even in the current digital computing era [3].

In the modern era, mathematics is no longer viewed solely as an abstract discipline but has evolved into an analytical tool capable of solving real-world problems through integration with information technology and computer science [4]. One outcome of this development is the K-Means algorithm, a data clustering technique that operates iteratively based on similarity of characteristics [5]. Although classified as a modern algorithm, K-Means has a logical structure that fundamentally reflects the classical algorithmic principles introduced by Al-Khawarizmi [6].

Previous studies have explored the application of the K-Means algorithm in data clustering, such as mapping COVID-19 cases,[7] clustering students in the SIAKAD system, and consumer segmentation.[8] However, to date, no explicit study has been found linking Al-Khawarizmi's algorithmic principles to the working structure of the K-Means algorithm, particularly in the context of modern data analysis. The absence of such a study indicates an important scientific gap that needs to be filled. Therefore, this study aims to analyse the contribution of Al-Khawarizmi's algorithmic principles namely systematic-logical, structural simplification, applied rationality, and quantitative approach to the structure of the K-Means algorithm, through a case study of household electricity access data in Indonesia in 2024.

By tracing the historical connection between classical Islamic scientific heritage and contemporary computational methods, this study is expected to not only provide new insights into algorithmic structures but also strengthen recognition of cross-civilisational contributions to the development of modern science. The emphasis on integrating classical epistemological values into digital analytical practices opens up new reflections on the importance of a transdisciplinary approach that combines values, logic, and technology.

2 Method

This study uses a quantitative approach with algorithmic analysis methods. The data used is secondary data in aggregate form, namely household electricity access data per province in Indonesia. This data was obtained from credible official sources and has been published, so there is no need for primary data collection or additional grouping based on regional categories. The selection of this data is based on its spatial complexity, which is relevant for testing the effectiveness of algorithmic principles in mapping distribution gaps. The research process began with a literature review to identify and formulate Al-Khawarizmi's main principles, such as systematic-logical, structural simplification, applied rationality, and quantitative-logical approach. These principles were conceptually operationalised within the K -Means framework through mapping to each stage of the algorithm: centroid initialisation (as initial variable

identification), data assignment based on Euclidean distance (as a form of structural simplification), centroid updating (solution evaluation), and iteration until convergence (as a form of applied rationality).

After the conceptual framework was established, the K-Means algorithm was applied to normalised electricity access data. The analysis included data pre-processing, determining the optimal number of clusters (K) through initial exploration and the elbow method principle, and iterative clustering based on the similarity of electricity access values. The clustering results were evaluated using quantitative metrics such as intra-cluster variance and inter-cluster distance. The validity of the research was strengthened through the selection of official data and analysis consistent with reputable scientific writing standards, as well as transparent explanations of each stage to ensure academic replicability and accountability. The interpretation of results also considers the classical Islamic philosophical values underlying Al-Khawarizmi's principles, as an integrative effort between the scientific heritage of civilisation and contemporary data analysis practices.

3 Result and Discussion

Al-Khawarizmi is a central figure in classical Islamic scholarship who, epistemologically, laid the foundations of mathematical rationality through a systematic approach to solving quantitative problems. His most renowned work, *Al-Kitāb al-Mukhtaṣar fī Ḥisāb al-Jabr wal-Muqābalah*, stands as a milestone in the consolidation of the discipline of algebra and also represents the initial formulation of algorithmic structures based on procedural and deterministic logic [9]. In the context of contemporary data science and computational technology, Al-Khawarizmi's scientific legacy demonstrates significant transhistorical relevance. The systematic, iterative, and applied patterns of thinking he formulated not only underpin the formation of modern programming logic but can also be traced in machine learning algorithms such as K-Means. Therefore, this discussion aims to construct a structural connection between Al-Khawarizmi's scientific principles and the working mechanism of K-Means, as well as to explore how this historical-conceptual integration can enrich our understanding of clustering algorithms in modern data analysis practices.

An algorithm is a series of steps or instructions arranged logically, systematically, and sequentially to solve a problem or perform a specific task. Algorithms serve as an effective method for performing calculations, data processing, decision-making, and problem-solving, both in computational systems and in everyday life [10]. The essence of an algorithm lies in the process of transforming input into desired output through a structured procedure, thereby producing accurate, efficient, and replicable solutions. Although the term 'algorithm' was only widely adopted in the modern computing era, the fundamental principles underlying it were formulated as early as the classical scientific era by Al-Khawarizmi, a Muslim scientist known for his rational, systematic, and practical approach to solving mathematical problems. Although conceptual and philosophical in nature, Al-Khawarizmi's fundamental principles can serve as an

epistemological foundation for contemporary algorithmic structures, including the K-Means algorithm as the latest innovation in the integration of mathematics and computing, ensuring that the integration of past knowledge remains relevant to the present.

Al-Khawarizmi's basic principles in algorithm development reflect a synthesis between mathematical rationality and applied values within the framework of classical Islamic epistemology. Conceptually, the four main principles that form the foundation of his thinking are: (1) logical systematicity, (2) structural simplification, (3) applied rationality, and (4) quantitative-logical approach[11]. The first principle of Al-Khawarizmi, logical systematicity, emphasises the importance of deterministic and replicable instruction sequences in solving problems. As Knuth (1997) emphasises, effective algorithms are always characterised by clear procedural regularity. This aligns with Al-Khawarizmi's method in *Al-Jabr wal-Muqabalah*, where solutions to equations are structured hierarchically, beginning with variable identification, form transformation, and culminating in final resolution. This approach demonstrates that procedural logic is not merely a technical tool but a scientific framework underlying problem-solving.

The second principle is structural simplification, realised through the concepts of *al-jabr* (recovery) and *al-muqabalah* (balancing), which can be examined in the context of symbolic manipulation in modern algorithm theory.[12] *Al-jabr* plays a role in eliminating negative terms through algebraic completion, while *al-muqabalah* reflects mathematical equivalence logic, which eliminates redundancy to simplify the model. These two principles serve as the initial prototype for mathematical transformations now used in algorithmic optimisation and simplifying computational complexity [13]. The next principle is applied rationality, which reflects the practical orientation of algorithm construction. Al-Khawarizmi believed that algorithms should be applicable in social life, such as inheritance calculations, trade ratios, and zakat calculations. This reflects the concept of instrumental rationality as explained by Weber, which involves the use of calculative logic to achieve efficiency in social actions.

Al-Khawarizmi demonstrated that he had already developed a form of applied rationality in mathematics that was inseparable from ethical values and societal needs. The final principle is the quantitative-logical approach, which reflects Al-Khawarizmi's orientation toward numerical relationships as real entities with causal connections in the physical world. Al-Khawarizmi did not view numbers as purely abstract entities but as representations of cosmic order. This approach demonstrates that mathematical processes, according to Al-Khawarizmi, are always supported by a strong deductive inference system that is rationally accountable. Al-Khawarizmi laid the foundation for the development of algorithms and also formed a scientific paradigm that unites logical structure, social function, and philosophical meaning [14]. This legacy became the initial construction of a transdisciplinary algorithmic thinking system that is relevant in the framework of contemporary science.

Table 1. Basic Principles and General Explanation

Basic Principles	General Explanation
Systematic & Step-by-Step	Developing logical and structured steps for problem solving
Simplification (<i>Al-Jabr & Al-Muqabalah</i>)	Simplifying equations through completion and balancing
Rationalisation & Applicability	Linking theory with real-world applications (inheritance, trade, etc.)
Quantitative & Logical	Using logical and quantitative approaches to solve complex problems

One of the algorithms that is an innovation in the modern era is the K-Means algorithm. K-Means is a clustering algorithm that falls under the category of unsupervised learning and is widely used in the fields of machine learning and data exploration [15]. The primary objective of this algorithm is to group data into k clusters based on similarity levels, ensuring that data within a single cluster share similar characteristics while clusters differ significantly from one another. This process is performed without requiring initial labels, making it an ideal method for identifying hidden patterns or structures in complex high-dimensional datasets. K-Means is popular in various applied fields such as customer segmentation, geographical area classification, and social behaviour clustering.

Technically, K-Means consists of four main iterative steps: (1) random initialisation of centroids; (2) assigning each data point to the nearest centroid based on a distance metric, typically Euclidean distance; (3) updating centroids by calculating the average position of points within each cluster; and (4) repeating the process until convergence is achieved, i.e., when there are no further significant shifts between clusters. Despite its simplicity, the K-Means algorithmic structure enables high computational efficiency and scalability in processing large volumes of data, making it one of the foundational yet crucial algorithms in data-driven exploratory analysis. The use of quantitative metrics such as Euclidean distance in the similarity determination process underscores K-Means' character as a spatial and mathematical representation-based algorithm.

The algorithm's operation is rooted in numerical logic and deductive structure, where iterations are performed to minimise total variance within clusters (intra-cluster variance) and maximise separation between clusters (inter-cluster distance). The relevance of this algorithm continues to grow in the era of data-driven decision-making, particularly when used to analyse spatial-based socio-economic phenomena, such as in the context of equitable access to household electricity, health service distribution, and multidimensional poverty mapping. In addition to the core algorithmic steps, K-Means also requires a pre-condition in the form of a numerical representation process for the data to be grouped. Each object in the dataset must be transformed into an n -dimensional vector, where each dimension represents a relevant quantitative feature. This process is known as vectorisation and forms the foundation of all distance-based clustering operations. At the end of its iterative cycle, K-Means stops the process when convergence is achieved, i.e., when the centroid no longer undergoes significant changes or when the cluster composition stabilises. This stage is known as the stopping

criteria and represents the achievement of an optimal solution that is proportional and computationally efficient [16].

Although classified as a modern computation-based algorithm, K-Means exhibits a logical structure that aligns with classical algorithmic principles. Each stage, from centroid initialisation to convergence, applies deterministic, iterative, and numerically-based procedures. This structure indicates that K-Means is a concrete form of algorithmic thinking that emphasises efficiency, stability, and rationality in data processing [17]. The iterative model used represents an effort to simplify the complexity of the dataset through classification based on spatial proximity and feature values. This framework reflects an epistemological approach historically pioneered by Al-Khawarizmi, where mathematical problems are solved through systematic, logical, and measurable steps. Both K-Means and classical algebraic systems place procedures at the foundation of reasoning, emphasising the importance of data representation transformation in achieving proportional solutions. Considering these structural and conceptual parallels, the relevance of Al-Khawarizmi's principles can be re-examined in the context of modern algorithm development. Mapping the correspondence between classical principles and K-Means stages serves as an interpretive framework for examining the continuity of algorithmic logic across time.

The initial stage in Al-Khawarizmi's method begins with identifying the problem structure through variable separation and determining the basic mathematical form to be solved. This approach parallels the centroid initialisation process in the K-Means algorithm, which involves selecting initial centre points as a reference for clustering. Centroid initialisation is a crucial component in cluster structure formation as it can significantly influence the final outcome. Both demonstrate that the selection of representative starting points is the foundation for procedural effectiveness in algorithmic processes, both in classical and modern computational contexts. Al-Khawarizmi's contribution in representing arithmetic problems in symbolic algebraic form serves as an abstraction process that enables logical manipulation of numerical phenomena. This symbolic representation is analogous to the data vectorisation process in K-Means, where entities are reduced to n -dimensional coordinates to enable distance calculation and spatial clustering. According to Bishop (2006), vectorisation is a key process in machine learning because it transforms complex data into a mathematically computable structure. Thus, both classical and modern systems emphasise the importance of representation as a transformative stage in data-based problem solving.

Al-Khawarizmi introduced the principles of *al-jabr* (completion) and *al-muqabalah* (balancing) as techniques to simplify the structure of equations to make them easier to solve. This process has parallels with the assignment stage in K-Means, where data is grouped based on proximity to a specific centroid. This grouping is a form of simplifying data complexity to create a more organised structure that can be analysed in a focused manner. In a modern context, data simplification through clustering has become an essential approach in big data management. In other words, both classical algebra and the K-Means algorithm aim to simplify structures while maintaining their informative significance. The evaluation stage in solving equations by Al-Khawarizmi is carried out by checking the equality and coherence of each step that has been taken. This procedure can be compared to the re-centroid process in K-Means, which is the

updating of the centroid position based on the average of the points in the previously obtained cluster. This iteration is an evaluative form that allows the model to move towards an optimal solution. In line with this, MacQueen (1967) states that K-Means is an iteration-based algorithm for minimising a specific objective function, usually the total variance in clusters [18]. Both emphasise that solution improvement is a gradual process that depends on the validation of the structure against the desired results.

In algebraic solutions, Al-Khawarizmi always emphasised the importance of a simple final form that could be applied in real-life contexts such as calculating zakat or inheritance. In K-Means, a similar approach is seen in the stopping criteria mechanism, which is the condition when iterations are stopped because the cluster structure has stabilised or the centroid no longer changes significantly. This convergence principle ensures that the final solution is an optimal form that is computationally efficient. In machine learning, stopping criteria serve as an indicator of the success of the learning process without overfitting. Convergence in K-Means and the final form in algebraic systems demonstrate that logical stability is the primary indicator of algorithmic success, both in classical and digital eras.

The structural relevance between Al-Khawarizmi's algorithmic principles and the stages of K-Means work emphasises that the logical and systematic foundations formulated in classical Islamic scientific tradition have conceptual continuity with modern computational methods. Every element in the K-Means algorithm, from initialisation to convergence, demonstrates procedural similarities and also reflects the deductive, quantitative, and applied approach to thinking as formulated by Al-Khawarizmi within the classical algebraic framework. This cross-temporal integration demonstrates that algorithms are products of epistemology that emerge through the dialectic between practical needs and rational reflection. Strengthening the connection between classical algorithmic heritage and contemporary clustering techniques like K-Means can open up new historical space in computational studies and offer an integrative paradigm capable of linking mathematical values, inferential logic, and data-based applications in a transdisciplinary manner.

The compatibility between the logical structure of K-Means and the fundamental principles of Al-Khawarizmi is further tested through direct application to household electricity access data in Indonesia. The dataset represents complex spatial conditions and requires a classification approach capable of capturing distribution patterns based on numerical similarity. The aim is to group regions based on electricity accessibility levels, enabling distribution patterns to be identified quantitatively and visually. The analysed data can be viewed in the following table:

Table 2. Percentage of Households by Province and Main Source of Lighting from Electricity (Percent)

Province	Percentage
Aceh	99.86
Northern Sumatera	99.73
West Sumatera	99.39
Riau	99.52

Jambi	99.58
South Sumatera	99.61
Bengkulu	99.88
Lampung	99.94
Bangka Belitung Island	99.87
Riau Island	99.91
Jakarta	100
West Java	99.99
Central Java	99.99
DI Yogyakarta	100
East Java	100
Banten	99.84
Bali	99.96
West Nusa Tenggara	99.98
East Nusa Tenggara	96.82
West Kalimantan	99.08
Central Kalimantan	99.26
South Kalimantan	99.92
East Kalimantan	99.98
Northern Kalimantan	99.49
Northern Sulawesi	99.87
Central Sulawesi	99.27
South Sulawesi	99.73
Southeast Sulawesi	99.89
Gorontalo	99.7
West Sulawesi	99.61
Maluku	97.25
Northern Maluku	98.72
West papua	96.3
Southwest papua	98.05
Papua	96.06
South Papua	80.61
Central Papua	56.08
Papua Mountains	73.66

After the household electricity access data was collected and normalised, the clustering process was conducted using the K-Means algorithm through SPSS software. This stage aims to group regions based on similarities in electricity access levels, so as to obtain a more targeted segmentation in reading the pattern of energy disparities in Indonesia. The number of clusters was determined as three ($k = 3$), by referring to the elbow method principle and exploratory analysis of the data distribution.

Table 3. Initial Cluster Centers

	Cluster		
	1	2	3

These values represent the initial position of the cluster centres based on the number of households with access to electricity. Cluster 1 indicates areas with very high coverage, Cluster 2 is in the medium category, and Cluster 3 represents areas with very low levels of electricity affordability. Philosophically, this stage can be linked to the principle of initial problem identification in Al-Khwarizmi's thought. In the classical algorithmic tradition, the first step of establishing the basic structure of the problem is the foundation for the entire solution process. The centroid initialisation in K-Means epistemologically reflects the effort to determine the main variables and initial conditions that become the foundation for the next iterative process, just as Al-Khwarizmi put variables and constants as the starting point in the preparation of algebraic equations. Thus, the selection of the initial centre is not only technical, but also reflects the continuity of Al-Khwarizmi's systematic logic in that every initial decision has logical implications for the stability and accuracy of the final result.

After initialisation, the K-Means algorithm runs an iterative process that aims to update the cluster centres until no significant change occurs.

Table 4. Iteration History

Iteration	Change in Cluster Centers		
	1	2	3
1	123.563	879.000	352.000
2	58.530	524.667	.000
3	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 3. The minimum distance between initial centers is 4391.000.

The results show that the algorithm converged at the third iteration, marked by no significant changes in the cluster centre coordinates (centroid). This indicates that the cluster structure has stabilised and is no longer undergoing significant shifts. Technically, this process reflects the algorithm's success in finding the optimal representation for each data group. From an epistemological perspective, this iterative stage is highly relevant to the process of evaluating solutions in Al-Khwarizmi's thinking system. Just as he emphasised the importance of re-testing each step of algebraic solution through repeated balancing and adjustment, the K-Means algorithm also takes a similar approach through centroid updates until numerical stability is achieved. Iteration is not only a technical mechanism but also a symbol of the deductive-evaluative approach to achieving harmony between data structure and cluster results. Furthermore, the achievement of convergence in the third iteration demonstrates the effectiveness of the initial structure that has been established. This

aligns with Al-Khawarizmi's principle of applied rationality, which holds that an ideal mathematical solution must not only be logically valid but also efficient and practically applicable. In other words, the stability achieved in the third iteration is not merely a statistical output but also reflects the success of the algorithmic logic in achieving an optimal and proportional final form. After the iterative process is complete, the final cluster centres are obtained, serving as the numerical representation of each data group.

Table 5. Final Cluster Centers

	Cluster		
	1	2	3
VAR0000 2	9933.97	7011.67	452.00

Based on the variable percentage of households with electricity as their main source of lighting (VAR00002), the final results of the clustering process using the K-Means algorithm show three cluster centres (final cluster centres), each with values of 9933.97 (Cluster 1), 7011.67 (Cluster 2), and 452.00 (Cluster 3). These values represent the average household electricity access in each group formed after the iteration process was completed and convergence was achieved. The significant difference between the centre values of Cluster 1 and Cluster 3 indicates a significant gap in the distribution of electricity access between regions. This demonstrates how the K-Means method can group data into a more informative and analytical structure, thereby quantitatively revealing spatial disparities. Conceptually, this final result can be interpreted as a form of ‘final resolution’ within the Al-Khawarizmi algorithmic framework, i.e., the achievement of a solution after undergoing a series of evaluative and transformational procedures. As in classical Al-Khawarizmi algebra, the final form of an equation must be simple, stable, and applicable in practical contexts. Thus, the final cluster centres in K-Means function as representative points of each group that can be used for more proportional data-based decision-making. This is also in line with the principle of stopping criteria, which indicates that the system has reached an optimal form and does not require further improvement.

The determination of these cluster centres is a manifestation of layered procedural logic. This is consistent with Al-Khawarizmi's approach to solving mathematical problems. Through deterministic and evaluative iterations, the K-Means algorithm demonstrates that each stage culminates in a single convergence point that philosophically resembles a stable and rational final solution in an algebraic model. Thus, the final cluster centres not only represent the technical output of the algorithm but also reflect the success of the method in applying the systematic, logical, and applicable principles inherited from the classical Islamic scientific tradition. To see the distribution of the population in each group, the clustering results show the following number of cases:

Table 6. Number of Cases in each Cluster

Cluster	1	31.000
	2	3.000
	3	5.000
Valid		39.000
Missing		.000

The distribution of cases across each cluster shows that out of a total of 39 valid data points, 31 belong to Cluster 1, 3 to Cluster 2, and 5 to Cluster 3. This composition reflects the dominance of certain characteristics that are widespread in the areas included in Cluster 1, which is the cluster with the highest average electricity access (9933.97), as obtained in the final cluster centres. The disparity in quantity between clusters also serves as an initial indicator of high spatial disparities in access to basic services such as electricity, especially when considered in the context of equitable development across regions. Algorithmically, these results demonstrate the success of K-Means in simplifying complex data structures into more internally homogeneous groups. This aligns with the principles of structural simplification (*al-jabr* and *al-muqabalah*) in Al-Khawarizmi's method, where entities that were originally diverse and scattered are condensed into categories that can be practically interpreted. These cluster categories not only facilitate analysis but also provide a strong foundation for evidence-based intervention policies.

To evaluate the quality and compactness of the clustering results, a number of quantitative validation metrics were used. First, the intra-cluster variance (variance within clusters) values were relatively low, especially in Cluster 1 and Cluster 3, indicating a high level of internal homogeneity within each group. Second, the inter-cluster distance measurements show a sharp separation between Cluster 1 and Cluster 3, indicating significant spatial disparities in electricity access between provinces in Indonesia. As further validation of cluster consistency, the average silhouette value was calculated using Euclidean distance as the similarity metric. The results showed a score of 0.592, which falls into the category of a fairly good cluster structure, where most provinces are classified accurately and have clear separation from other clusters. This metric reinforces the conclusion that the K-Means algorithm successfully captures hidden patterns in the distribution of household electricity access.

Philosophically, this evaluative stage aligns with the principles of applied rationality and evaluative iteration developed by Al-Khawarizmi, where each problem-solving stage is directed toward achieving an optimal, balanced, and socially meaningful solution. The use of this metric not only reinforces the technical validity of the algorithm but also provides a measurable basis for interpreting the classical epistemological legacy in the context of contemporary data analysis.

Furthermore, the differences in the number of members between clusters also show how the iterative K-Means model seeks a balance between data representation and classification efficiency. Within Al-Khawarizmi's algorithmic framework, this reflects applied rationality—that is, how the final results must be relevant to real-world needs and usable in social practice. Therefore, these results not only present quantitative data but also open up space for reflection on how modern algorithms can be used to bridge

inequality issues through a historically and transformatively scientific approach. Thus, the integration of the K-Means algorithm in the clustering of household electricity access data not only demonstrates the technical effectiveness of the clustering method but also provides philosophical reflection on the continuity of Al-Khawarizmi's algorithmic principles in modern computational practice. The combination of classical logical structures and contemporary data technology opens up new possibilities for building a transdisciplinary and historical-contextual scientific paradigm.

The results of the 2024 household electricity access data clustering in Indonesia using the K-Means algorithm with three clusters show segmentation patterns that reflect the quantitative level of electricity accessibility between regions. The first cluster, comprising 31 provinces, represents regions with an average electricity access rate of 99.33%. This indicates that most provinces in Indonesia have enjoyed equitable and stable electricity services, likely located in urban areas and regions with established energy infrastructure. The second cluster includes 3 provinces with an average electricity access rate of 70.11%. These regions are in the moderate category, indicating that although electricity services are available to the majority of the population, there are still groups of people who are not fully covered, either due to geographical, economic, or resource distribution constraints.

In addition to showing spatial segmentation in household electricity access in Indonesia, the clustering results also clearly reflect Al-Khawarizmi's algorithmic principles. Cluster 1, consisting of 31 provinces with an average electricity access rate of 99.33%, represents a category of regions with very stable energy infrastructure. Logically and systematically, the provinces in this cluster meet the criteria of high homogeneity and uniform numerical representation—in line with Al-Khawarizmi's systematic-logical principles, which emphasise the importance of order in the structure of solutions. This classification reflects the process of identifying the basic form of the problem, as in the *al-jabr wal-muqabalah* method, where dominant elements are simplified into representative patterns.

Cluster 2, comprising 3 provinces with an average of 70.11%, is in a transitional position. This cluster describes regions with uneven energy achievements but with potential for improvement. In principle, this reflects Al-Khawarizmi's evaluative approach—where a problem is analysed step by step through iterations until a proportional solution is reached. This cluster serves as a logical space for considering structural differences, while also reflecting the value of applied rationality, where temporary solutions still have practical functions in a social context. Meanwhile, Cluster 3, with 5 provinces and an average electricity access rate of only 4.52%, represents regions that are lagging behind and marginalised in the context of energy development. The existence of this cluster highlights the importance of Al-Khawarizmi's quantitative-logical principles in understanding seemingly extreme entities as integral parts of the system's structure. This regional classification is a concrete form of structural simplification through data-based categorisation, emphasising the readability of information within a complex system.

Meanwhile, the third cluster contains 5 provinces with an average access rate of only 4.52 percent, reflecting very significant limitations in electricity access. The regions in this cluster are likely to be 3T (underdeveloped, frontier, and outermost)

areas that still experience gaps in energy infrastructure development. Overall, this segmentation confirms the existence of spatial disparities in energy distribution, even though national electricity access is already high. The clustering process successfully grouped provinces into a more organised structure based on proximity of values, resulting in a clearer map of disparity distribution. Conceptually, these results align with the principles of structural and evaluative simplification within the Al-Khawarizmi algorithmic framework, which prioritises a systematic and logical approach to reducing complexity towards a more stable and applicable final form. These results can serve as a basis for developing more targeted energy distribution strategies. By combining classical algorithmic logic with modern technology, this analysis offers rational and contextually appropriate policy alternatives.

The segmentation results obtained through the K-Means algorithm are not only analytically important, but also have significant practical implications for the formulation of data-driven energy distribution strategies. Each cluster represents a group of regions with vastly different electricity access conditions: regions with established energy infrastructure (Cluster 1), transitional areas with moderate accessibility (Cluster 2), and regions still experiencing extreme electricity access limitations (Cluster 3). This pattern can be leveraged by policymakers to design more targeted and needs-based interventions, with proportional resource allocation.

Furthermore, this classification approach is in line with the principles of systematic and evaluative logic inherited from Al-Khawarizmi. Energy distribution strategies based on the results of clustering are not only based on statistical data, but also represent the actualisation of rational and structured solution development—starting from problem identification, simplification of complexity, to the determination of a stable final form. Thus, the application of these clustering results can support more equitable, efficient, and philosophical energy policies, in line with the spirit of distributive justice ('*adl*) in classical Islamic science.

4 Conclusion

This study successfully demonstrated that the algorithmic principles formulated by Al-Khawarizmi, such as systematic, logical, and quantitative approaches, not only have historical value but can also be integrated in a relevant manner into modern analytical methods such as the K-Means algorithm. Through a case study of household electricity access in 39 provinces in Indonesia, the clustering results revealed a structured segmentation pattern: groups with very high electricity access (Cluster 1), moderate access (Cluster 2), and very low access (Cluster 3). This pattern not only illustrates spatial inequality but also represents a classification process aligned with the scientific logic inherited from Al-Khawarizmi, namely simplifying complexity through iterative and evaluative structures.

However, the primary strength of this research lies not only in its theoretical contributions. These findings open up concrete opportunities for the application of algorithms informed by values of fairness and rationality in the formulation of energy

distribution policies. For example, the results of this clustering can serve as a basis for developing more proportional strategies for the distribution of electrical infrastructure, where provinces in Cluster 3 are prioritised for intervention. Furthermore, this ethical algorithmic principle can be adopted in the development of decision-making systems in other public sectors, such as educational infrastructure planning, social assistance management, and digital equity policies. This approach has the potential to shape data processing and policy-making systems that are not only technically efficient but also socially and contextually fair.

However, this research also has limitations, particularly in terms of policy validation and the impact of implementation in the field. The model developed is still analytical and based on secondary data, so it has not yet tested how the cluster results can influence the formulation of actual policies. Therefore, further steps are needed in the form of implementative studies, cluster-based policy simulation tests, and interdisciplinary collaboration between data scientists, development planners, and policy makers. Thus, the algorithmic approach inspired by classical epistemology is not only a philosophical reflection but also transforms into a practical instrument in addressing the challenges of equitable development in the digital age.

In addition to producing clear segmentation of household electricity access patterns in Indonesia, the use of the K-Means algorithm also reflects structural compatibility with Al-Khawarizmi's algorithmic principles. The iterative stages of K-Means—from initialisation to convergence illustrate a systematic, logical, and evaluative approach, akin to the classical algebraic approach that is systematically constructed towards a stable and applicable solution. This deterministic structure demonstrates that modern algorithms like K-Means are not merely technical classification tools but also representations of the legacy of classical Islamic scientific logic, which emphasises procedural regularity and efficiency in problem-solving. As a comparison, alternative methods such as hierarchical clustering could be used in future research to evaluate whether similar classification structures can be generated through a non-iterative, relation-based approach. While visually appealing and flexible, hierarchical structures do not directly reflect Al-Khawarizmi's algorithmic framework, which is oriented toward efficient convergence and applied rationality. Thus, the selection of K-Means in this study is not only based on computational practicality but also on its philosophical value, making it more representative in bridging classical science and contemporary analytical practices.

References

- [1] E. L. Khalil, "Does the state usher in a special stage in history? Probing *the Dawn of Everything: A new history of humanity*," *Australian J of Anthropology*, hlm. taja.70012, Jun 2025, doi: 10.1111/taja.70012.
- [2] A. Nawallia dan N. Mishriya, "Kontribusi Ilmuwan Muslim dalam Pembentukan Matematika dan Perkembangan Matematika dalam Sejarah Peradaban Islam," vol. 3, no. 2, 2024.

- [3] G. Avazbek, "MUHAMMAD IBN MUSA AL-KHWARIZMI REPRESENTATIVE OF THE ISLAMIC RENAISSANCE (783-850)," vol. 3, no. 5, 2025.
- [4] E. Mailani, N. Rarastika, D. A. Purba, I. Novita, dan J. R. B. Ginting, "ANALISIS PENERAPAN DEEP LEARNING DALAM PEMECAHAN MASALAH MATEMATIKA SISWA SEKOLAH DASAR," vol. 3.
- [5] M. Ahmed, R. Seraj, dan S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics*, vol. 9, no. 8, hlm. 1295, Agu 2020, doi: 10.3390/electronics9081295.
- [6] Y. Liang *dkk.*, "Foundation Models for Spatio-Temporal Data Science: A Tutorial and Survey," 12 Maret 2025, *arXiv*: arXiv:2503.13502. doi: 10.48550/arXiv.2503.13502.
- [7] D. N. P. Sari dan Y. Sukestiyarno, "Analisis Cluster dengan Metode K-Means pada Persebaran Kasus COVID-19 Berdasarkan Provinsi di Indonesia," vol. 4, 2021.
- [8] K. S. Amalia, A. Purbaningrum, dan R. Kusumastuti, "Analisis Pengelompokan Data Karyawan Terbaik Perusahaan Menggunakan Algoritma K-Means Clustering," 2023.
- [9] S. N. S. Sahuri, "Motivasi Saintis Muslim Dalam Perkembangan Ilmu STEM: Sorotan Literatur Bersistematik," vol. 10, no. 1, 2025.
- [10] Z. M. Nasution, M. M. Siadari, I. J. S. Saragih, I. O. Kirana, dan Z. A. Siregar, "Penerapan Matematika Algoritma dalam Bidang Komputer," *F.JMPM*, vol. 6, no. 2, hlm. 180–191, Des 2023, doi: 10.47662/farabi.v6i2.634.
- [11] Putri Wulandari, Samsu Alam.L, Yulia Afriliana, Jamrizal Jamrizal, dan Samsu Samsu, "Peran Kepemimpinan Rasulullah dalam Pengembangan Pendidikan Islam dan Warisan Peradaban Islam," *Hikmah*, vol. 2, no. 1, hlm. 20–34, Des 2024, doi: 10.61132/hikmah.v2i1.512.
- [12] S. S. Alam, "MATHEMATICS FOR ALL AND FOREVER," 2015.
- [13] M. Baharuddin dan W. S. W. Abdullah, "THE BOOK AL-JABR WA AL-MUQĀBALAH: A RESEARCH ON ITS CONTENT, WRITING METHODOLOGY AND ELEMENTARY ALGEBRA BY AL-KHWĀRIZMĪ," 2014.
- [14] "ISLAM-DAN-KONTRIBUSINYA-TERHADAP-PERKEMBANGAN-ILMU-PENGETAHUAN-SUATU-TELAHAH.pdf."
- [15] S. Naeem, A. Ali, S. Anam, dan M. M. Ahmed, "An Unsupervised Machine Learning Algorithms: Comprehensive Review," *IJCDS*, vol. 13, no. 1, hlm. 911–921, Apr 2023, doi: 10.12785/ijcnds/130172.
- [16] A. A. Wani, "Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions," *PeerJ Computer Science*, vol. 10, hlm. e2286, Agu 2024, doi: 10.7717/peerj-cs.2286.
- [17] M. Ahmed, R. Seraj, dan S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics*, vol. 9, no. 8, hlm. 1295, Agu 2020, doi: 10.3390/electronics9081295.
- [18] C. Holder dan A. Bagnall, "Rock the KASBA: Blazingly Fast and Accurate Time Series Clustering," 26 November 2024, *arXiv*: arXiv:2411.17838. doi: 10.48550/arXiv.2411.17838.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

