International Journal of Islamic Economics (IJIE)
P-ISSN: 2686-2131 E-ISSN: 2686-2166
DOI: https://doi.org/10.32332/ijie.v8i02.13481 Vol 8, No 2, (2026) Pages. 280-294

Machine Learning-Based Clustering for Optimizing Zakat, Infaq, and Sadaqah Fund Distribution Among Ahmad Dahlan University Students

Safika Maranti[1], Sugiyarto Surono[2], Nurhidayati Kasrul Jalil[3], Khusnul Khotimah[4], Sintia Afriyani[5]
Universitas Ahmad Dahlan, Yogyakarta, Indonesia[1,2]
Sakarya University of Turkey, Turkey[3]
Universitas Islam Indonesia, Yogyakarta, Indonesia[4,5]

Article History:	Abstract
Corresponding Author: Safika Maranti safika.maranti@lpsi.uad.ac.id **Submitted:** April 30th, 2026 **Revised:** July 6th, 2026 **Accepted:** August 16th, 2026 **Published:** September 23rd, 2026 	**Objective:** This study aims to identify latent patterns in the distribution of zakat, infaq, and sadaqah funds among student beneficiaries at Universitas Ahmad Dahlan by grouping recipients according to their socio-economic, academic, and social characteristics. **Method:** This study adopts a quantitative approach using unsupervised machine learning techniques. The K-Means and Density-Based Spatial Clustering of Applications with Noise algorithms were applied to secondary data from Lazismu Ahmad Dahlan University for 2023–2026. After cleaning, the study retained 185 valid recipient records from an initial dataset of 194 observations. Data preprocessing included encoding and normalization. The study compares centroid-based (K-Means) and Density-Based Spatial Clustering of Applications with Noise clustering methods to identify recipient segments. **Results:** The findings show that Density-Based Spatial Clustering of Applications with Noise produces more compact and better-separated clusters than K-Means. The algorithm also identifies atypical recipients as noise, revealing unique beneficiary profiles. The results indicate substantial heterogeneity among recipients. Some economically vulnerable students demonstrate strong academic performance yet receive relatively lower assistance, while other groups receive higher levels of support despite less vulnerable conditions. **Implications:** The findings suggest that clustering analysis can support more objective, transparent, and adaptive ZIS fund distribution. Integrating such analysis into planning, implementation, and evaluation processes may improve targeting accuracy and promote equitable, evidence-based decision-making while still considering individual circumstances. **Keywords:** Clustering; Density-Based Spatial Clustering; Islamic Social Finance; K-Means; Machine Learning; Zakat Infaq Sadaqah.

Introduction

Zakat, infaq, and ṣadaqah (ZIS) are strategic instruments in poverty alleviation and improving access to education, especially for economically disadvantaged groups. (Faiza et al., 2023; Nabila, 2022). From an Islamic perspective, zakat is not only an individual religious obligation but also has an important socio-economic function in promoting social justice, reducing inequality, and improving community welfare. (Baidhowi & Triwibowo, 2023; Bawana et al., 2023; Rahmawati & Yazid, 2025). The use of ZIS funds in the education sector is in line with the objectives of maqasid al-shari'ah, especially in preserving and developing the intellect (hifz al-aql). Several contemporary studies show that distributing zakat for education has a sustainable effect because it can increase individual capacity and encourage social mobility.

Education plays a fundamental role in improving the quality of human resources, opening up employment opportunities, and reducing socioeconomic inequality. (Almaida, 2023). Therefore, every citizen has the right to obtain a proper education (Iskandar, 2019) as stipulated in Article 28C of the 1945 Constitution of the Republic of Indonesia: "Every person shall have the right to develop himself/herself through the fulfillment of his/her basic needs, shall have the right to education and to benefit from science and technology, arts and culture, in order to improve his/her quality of life and for the welfare of humankind." However, in reality, finances remain a major obstacle for some students in completing their studies. In this context, the use of ZIS funds to support the continuity of student studies not only represents a religious obligation, but also has social and economic dimensions that have a long-term impact (Mohammad & Maulida, 2020).

In higher education institutions, including Ahmad Dahlan University (UAD), ZIS funds play an important role in helping students facing financial difficulties continue their studies. Lazismu UAD manages ZIS fund distribution at UAD as the campus zakat institution. In its operations, the selection mechanism for aid recipients is still dominated by administrative criteria and subjective recommendations, which has the potential to cause bias, inefficiency, and a lack of accuracy in targeting (Fakiriyah et al., 2023; Saiful et al., 2025). This situation becomes increasingly challenging as the number of aid applicants increases and the socioeconomic backgrounds of students become more diverse (Rahmawati & Yazid, 2025) (Rahmadi & Mu'minin, 2024). In fact, student data on family economic aspects, academic achievements, and social backgrounds is available and can support more objective, data-driven decision-making.

Despite the growing number of studies on ZIS fund management, most previous research has focused on recipient eligibility assessment, decision support systems, and classification models to determine aid recipients (Hastuti, 2014; Nurrahman et al., 2025; Pratama, 2023; Triayudi, 2023). While these studies have improved administrative efficiency, they generally rely on predefined categories and offer limited insight into latent patterns among beneficiaries. Consequently, a research gap remains in identifying the natural heterogeneity of ZIS recipients based on their socio-economic, academic, and social characteristics, particularly in higher education institutions.

To address this gap, this study applies an unsupervised machine learning approach using K-Means and DBSCAN clustering algorithms to explore the underlying structure of recipient characteristics without predefined labels. This study's novelty lies in three aspects. First, it applies clustering techniques to analyze ZIS recipient profiles in a university setting, which remains relatively underexplored in Islamic social finance research. Second, it compares centroid-based clustering (K-Means) and density-based clustering (DBSCAN) to evaluate their effectiveness in identifying recipient segments. Third, it integrates machine learning analysis with the principles of Islamic social finance and *maqasid al-shari'ah*, providing a data-driven perspective to support more transparent, equitable, and accountable ZIS fund distribution.

Literature Review

Previous Study

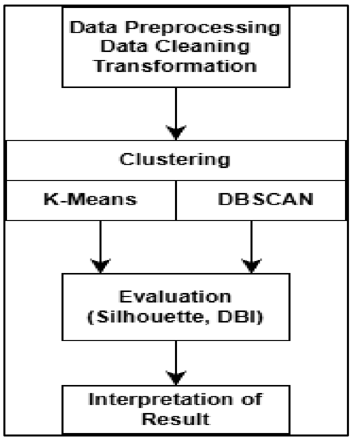
Alongside technological developments that open new opportunities in Islamic social fund management, recent studies have examined the application of machine learning methods to improve the effectiveness of ZIS management. (Alwi et al., 2023). For example, a decision support system (DSS) uses a combination of the Random Forest and Fuzzy Analytical Hierarchy Process (F-AHP) methods, where Random Forest is used to determine the priority level of recipients (Nurrahman et al., 2025). In addition, the K-Nearest Neighbor method is also used to group potential *mustahiq* so that distribution is more targeted (Triayudi, 2023). Other methods, such as Naïve Bayes, which has considerable potential, are also used to examine the eligibility of *mustahiq* (Aprianto et al., 2025). Previous studies have shown that machine learning can support decision-making in distributing ZIS funds. However, most studies focus on labeled-data classification, while this study uses data on ZIS recipients who have already received funds to identify patterns and characteristics. Thus, this study is novel in applying an unsupervised learning approach, specifically clustering, to analyze ZIS fund recipients based on the similarity of student characteristics at UAD. This approach groups aid recipients by attribute similarity without requiring class labels or assumptions about the number of groups at the outset. By using ZIS recipient data on economic conditions, academic achievement, and social background, the clustering method is expected to reveal the natural structure of aid recipients, such as groups of students with very low economic conditions, middle-income groups, and high-achieving students with financial limitations. In addition to machine learning classification approaches, studies have also explored clustering and data mining techniques for zakat analysis. The K-Means algorithm has been widely discussed as an efficient clustering technique for grouping data based on similarity patterns (Ahmed et al., 2020). In the context of zakat management, clustering has also been applied to identify regional variations in zakat needs and support more strategic allocation (Efita & Triase, 2024). Similarly, other work on k-means clustering for the distribution of zakat indicates the potential of unsupervised learning in classifying zakat recipients based on multiple attributes without predefined class labels (Islami et al., 2023). Furthermore, beyond pure machine learning, the implementation of data mining using Naïve Bayes at BAZNAS East Lombok demonstrated high classification accuracy for identifying eligible *mustahik*, indicating its practical applicability in real institutional settings (Ismail & Alsadhan, 2023).

Research Framework

This study uses a descriptive quantitative approach with clustering methods to group recipients of zakat, infaq, and sadaqah (ZIS) funds based on similarities in socioeconomic and academic characteristics. This approach was chosen to explore natural patterns in ZIS recipient data without prior class labels, aligning with unsupervised learning (Islami et al., 2023). The two main algorithms used in this study are K-Means and DBSCAN. K-Means was chosen for its ability to form homogeneous clusters with high computational efficiency (Hendrastuty, 2024), while

DBSCAN was used to detect density-based patterns and identify natural clusters and anomalous data without requiring the number of clusters to be determined in advance (Aprianti et al., 2024). The entire analysis was carried out in Python using the pandas library for data processing (Simbolon & Friskila, 2024), scikit-learn to implement the K-Means and DBSCAN algorithms, and matplotlib to visualize the results. Figure 1 illustrates the study workflow.

Figure 1.
Workflow of the Clustering



Source image: Author

Research Methodology

The population in this study consists of all recipients of zakat, infaq, and sadaqah (ZIS) funds registered with the Muhammadiyah Zakat, Infaq, and Sadaqah Institution (Lazismu) at Ahmad Dahlan University (UAD) for the periods 2023–2024, 2024–2025, and 2025–2026. The initial dataset included 194 recipient records: 38 from 2023–2024, 59 from 2024–2025, and 97 from 2025–2026. After cleaning the data by removing incomplete, duplicate, or inconsistent records, 185 records were retained for clustering analysis. Because the data cover all fund recipients during these periods, the sampling technique used is total sampling, in which all recipient data are used as the unit of analysis. This approach is in line with the objective of unsupervised learning, which focuses on discovering intrinsic structures and hidden patterns from the complete dataset rather than making statistical inferences from a subset of observations (Saputri, 2025).

The data used in this study are secondary data obtained from the Ahmad Dahlan University Zakat, Infaq, and Sadaqah Institution (Lazismu UAD). The dataset includes information on ZIS fund recipients, including variables such as family economic conditions (as indicated by a certificate of financial hardship), Grade Point Average (GPA), non-academic achievements, organizational involvement, and the amount of financial assistance received. We selected these variables because they represent the socioeconomic, academic, and social dimensions relevant to evaluating ZIS fund distribution. We collected all data digitally and analyzed it to identify similarities in recipient characteristics.

Before cluster analysis, we performed several preprocessing steps. First, we cleaned the data to identify and remove duplicate records and handle missing values. We reviewed records with incomplete or inconsistent information and excluded them when necessary to maintain data quality. Second, we transformed categorical variables into numerical representations using encoding techniques to enable machine learning processing. Third, we normalized all numerical variables using Min-Max scaling to ensure all attributes were measured on a comparable scale and to prevent variables with larger numerical ranges from dominating the clustering process (Rao et al., 2023). We encoded the categorical variables, including SKTM ownership, non-academic achievement, and organizational involvement, as binary numerical values, where 1 indicates the presence of the attribute, and 0 indicates its absence. We normalized the nominal assistance variable and GPA using Min-Max Scaling. After cleaning, we excluded 9 records due to incomplete or inconsistent information, leaving 185 valid records for analysis. To validate the preprocessing stage, we re-examined the dataset to ensure no missing values, duplicate records, or inconsistent data formats remained before clustering analysis.

We used K-Means and DBSCAN for clustering. For K-Means, we tested several values of k to identify the most appropriate number of clusters. The elbow method determined the optimal k by analyzing the Within-Cluster Sum of Squares (WCSS) across different cluster configurations. The elbow point on the WCSS curve indicated that k = 5 provided the most suitable cluster structure, balancing cluster compactness and interpretability. (Taha et al., 2024). This approach is widely used in clustering studies to determine the number of clusters before implementing the model. For DBSCAN, the parameter values of epsilon (ε) and min_samples were determined through iterative experimentation and examination of the data distribution. These parameters were selected to balance identifying dense clusters while effectively detecting noise or atypical observations, a key advantage of DBSCAN. (Monko & Kimura, 2023). Using both algorithms allows comparison between centroid-based and density-based clustering approaches in identifying recipient group characteristics.

The clustering results were then evaluated using the Silhouette Score to measure intra-cluster cohesion and inter-cluster separation. (Chuang & Chou, 2025)as well as the Davies-Bouldin Index (DBI) to assess cluster quality based on the ratio between within-cluster similarity and between-cluster separation (Hasan, 2024) . Higher Silhouette Score values and lower DBI values indicate better clustering performance. These evaluation metrics formed the basis for selecting the most appropriate clustering results. After the evaluation, we interpreted each cluster based on the average values and dominant characteristics of its constituent attributes to identify recipient groups, such as students with low economic status, moderate economic status, high academic achievement, or distinctive social profiles. The results of this interpretation were subsequently used to formulate recommendations for ZIS fund distribution policies that are more targeted, transparent, and data-driven.

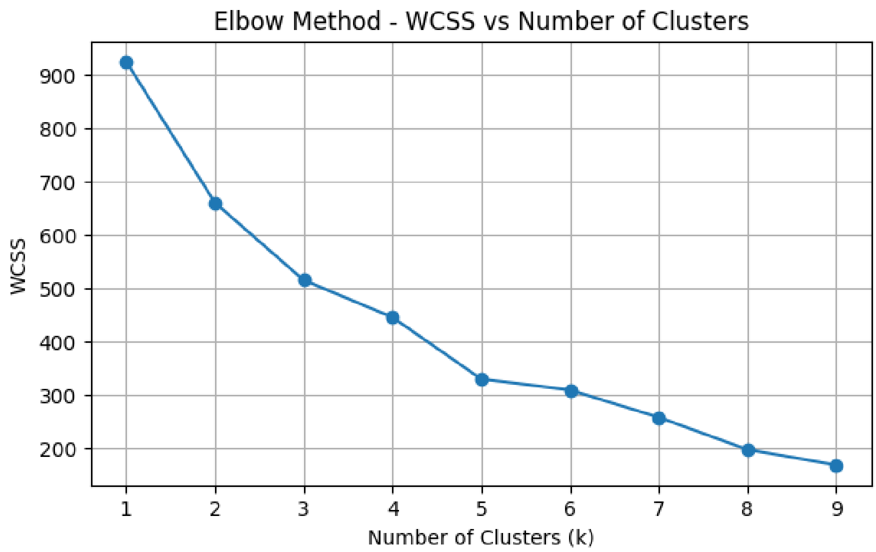
Results and Discussion

K-Means Clustering

In applying the K-Means algorithm, the first step is to determine the optimal number of clusters. This is done using the elbow method, which analyzes changes in the within-cluster sum of squares (WCSS) value in relation to the number of clusters. The elbow point on the graph indicates the most representative number of clusters, where adding more clusters no longer significantly decreases WCSS. Based on the elbow method results, this study used five clusters, as shown in Figure 2.

Figure 2.

Elbow Method for optimal value of k



Source image: Author

Figure 2: Using 5 clusters, 5 groups of ZIS fund recipients were generated. Table 1 summarizes the average value of each variable in each cluster. The variables analysed include possession of a Certificate of Poverty (SKTM), Grade Point Average (GPA), non-academic achievements, organizational activity, and the nominal amount of assistance received.

Table 1.

Average value of each variable for each cluster

Cluster	SKTM	GPA	Non-academic achievements	organizational activity	Nominal
0	1.00	3.74	0.00	1.00	3.49 million
1	0.00	3.76	0.58	1.00	4.59 million
2	1.00	3.79	1.00	1.00	2.99 million
3	0.44	3.76	0.37	0.00	4.33 million
4	0.95	3.28	0.90	1.00	3.29 million

Source table: Author

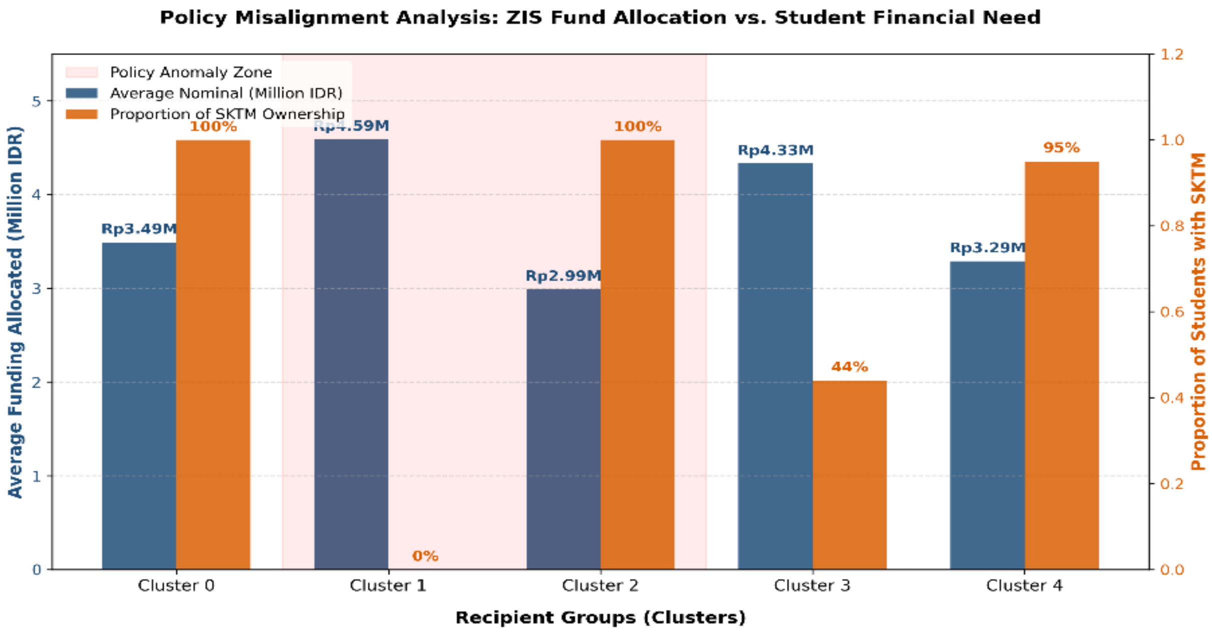
Cluster 0 is characterized by all members having SKTM, relatively high GPA (3.74), no non-academic achievements, all active in organizations, and receiving an average of Rp3.49 million in assistance. Cluster 1 consists of students without SKTM, high GPA (3.76), moderate achievements, active in organizations, and receiving the highest assistance, namely an average of Rp4.59 million. Cluster 2 consists of members who all have SKTM, the highest GPA (3.79), all have non-academic achievements and are active in organizations, but receive the lowest assistance, namely an average of Rp2.99 million. Cluster 3 has a low proportion of SKTM (44%), a high GPA (3.76), low achievements, is not active in organizations, and receives relatively high assistance of Rp4.33 million. Cluster 4 is dominated by students with SKTM (95%), the lowest GPA (3.28), high achievements, active in organizations, and receives average assistance of Rp3.29 million.

These cluster characteristics reveal substantial variation in the socio-economic and academic profiles of ZIS recipients. In particular, the contrast between Cluster 1, which receives the highest average assistance despite lacking SKTM ownership, and Cluster 2, which consists entirely of economically disadvantaged students yet receives the lowest average assistance, highlights patterns that warrant further examination. However, interpret these findings with caution. The dataset used in this study does not capture all qualitative considerations that may influence aid allocation decisions, such as urgent family circumstances, health-related issues, institutional recommendations, or other contextual factors considered by Lazismu UAD. Therefore, the clustering results should be viewed as indicators for further review and policy evaluation rather than as evidence of misallocation or inequitable distribution.

The quality of the clustering results was evaluated using the Silhouette Score and Davies–Bouldin Index (DBI). The K-Means model produced a Silhouette value of 0.3997 and a DBI value of 0.9999. Silhouette values in this range indicate that the objects in each cluster have a fairly good level of internal similarity, although there is still some overlap between clusters. Meanwhile, a DBI value close to 1 indicates moderate separation between clusters. Overall, these results show that the cluster structure is representative enough to serve as an exploratory tool for understanding the characteristic patterns of ZIS fund recipients.

The K-Means clustering results show significant heterogeneity in ZIS recipients' characteristics across economic, academic, and social activities. Specifically, Cluster 2 represents students who, by norm, have the potential to receive priority assistance because they are economically disadvantaged (have SKTM), have high academic and non-academic achievements, and are active in organizations, but receive the lowest assistance compared with other clusters. Conversely, Clusters 1 and 3 receive relatively more assistance even though not all their members are in the most economically vulnerable conditions.

Figure 3.
Policy Misalignment Analysis



Source image: Author

Figure 3 compares the average amount of ZIS assistance received by each cluster with the proportion of students holding a Certificate of Financial Hardship (SKTM). The figure provides an additional perspective on the relationship between economic vulnerability and funding allocation. Cluster 2 demonstrates a combination of high economic vulnerability and strong academic performance, yet receives the lowest average assistance. In contrast, Cluster 1 receives the highest average assistance despite lacking SKTM ownership. These findings suggest that factors beyond economic hardship alone may influence assistance allocation. This difference does not necessarily indicate an error in the distribution policy; rather, it reflects that a data-based approach can serve as a complementary instrument to provide additional perspective in the decision-making process. Thus, these clustering results could serve as an analytical tool to improve transparency, objectivity, and accountability in future planning for the distribution of ZIS funds.

From an Islamic social finance perspective, these findings raise important considerations regarding distributive justice and fairness. Although clustering results cannot determine whether allocation decisions are correct or incorrect, they can identify patterns that warrant further institutional evaluation. Students who demonstrate both economic vulnerability and academic excellence may represent priority groups for support under the broader objectives of maqasid al-shari'ah. Therefore, machine learning-based clustering should be viewed as a complementary decision-support tool that enhances transparency and accountability rather than replacing human judgment in ZIS management.

DBSCAN Clustering

In this study, we used the k-distance plot to determine the parameters ε (epsilon) and min_samples. We set min_samples to 7, considering the number of data dimensions used in the analysis (five variables) and following common practice in the literature that min_samples should be greater than or equal to the number of dimensions plus one so the clusters represent a stable density structure. Next, we determined ε from the k-distance plot; the elbow point was at ε = 0.45 with min_samples = 7. We calculated and sorted the distances to each point's seven closest neighbours, then plotted them to identify the elbow point indicating a significant change in the distance gradient. We selected this point as ε because it reflects the data's natural density limit, allowing more appropriate cluster separation in line with the characteristic distribution structure of ZIS fund recipients.

Table 2.
Cluster Size Distribution

Cluster Label	Category	Frequency	Percentage
-1	Noise	39	21.08%
0	Cluster 0	44	23.78%
1	Cluster 1	85	45.95%
2	Cluster 2	9	4.86%
3	Cluster 3	8	4.32%

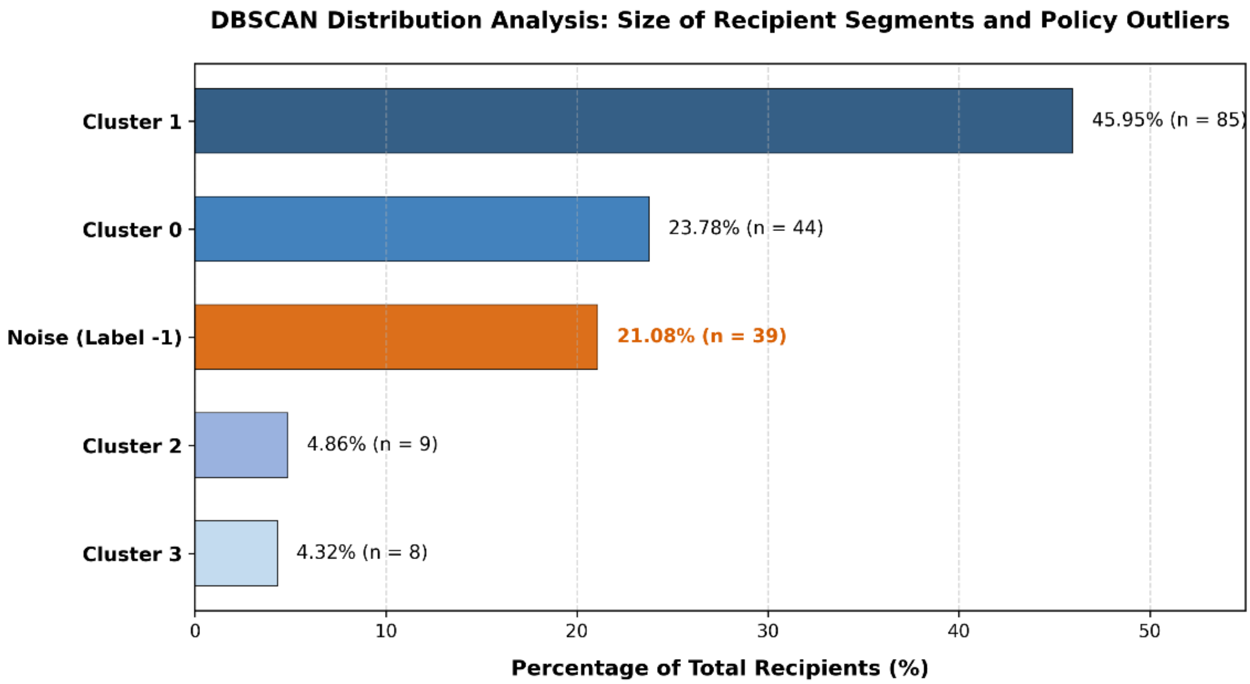
Source table: Author

Applying the DBSCAN algorithm produced four main clusters and a few data points categorized as noise (label -1). The distribution of members in each cluster shows that Cluster 1 is the largest, with 85 members, followed by Cluster 0 with 44 members, while Cluster 2 and Cluster 3 have only 9 and 8 members, respectively. We categorized 39 data points as noise, meaning they did not belong to any cluster because their characteristics differed significantly from the majority group.

We evaluated the quality of the DBSCAN clustering results using the Silhouette Score and Davies–Bouldin Index (DBI). The DBSCAN model produced a Silhouette value of 0.483 and a DBI value of 0.724 (calculated without including noise data). A Silhouette score near 0.5 indicates fairly good internal coherence and relatively clear separation between clusters. Meanwhile, a DBI value below 1 indicates that the clusters are compact and do not overlap excessively. This shows that the DBSCAN-derived cluster structure is stable and representative for exploratory analysis.

The dense clusters in the DBSCAN results indicate recipient groups with highly homogeneous characteristics, which can form the basis for stable policy segmentation. Meanwhile, the large proportion of noise data indicates that there are a few recipients with unique or extreme characteristics that general patterns cannot explain. This shows that the ZIS distribution system cannot be fully standardized and requires policy flexibility to handle special cases, such as students with severe economic hardship, unusual family circumstances, or specific urgent needs.

Figure 4.
DBSCAN Distribution Analysis



Source image: Author

Figure 4 illustrates the relative size of each cluster generated by DBSCAN. The visualization highlights that nearly half of the recipients belong to Cluster 1, while approximately 21% of observations are classified as noise. This relatively large proportion of noise indicates recipients whose characteristics differ substantially from dominant recipient profiles, suggesting considerable heterogeneity within the beneficiary population. Thus, the DBSCAN results complement the K-Means results by showing that, beyond general segmentation, ZIS management needs an individualized approach. Hence, distribution policies are not only statistically efficient but also sensitive to the diversity of recipients' social conditions.

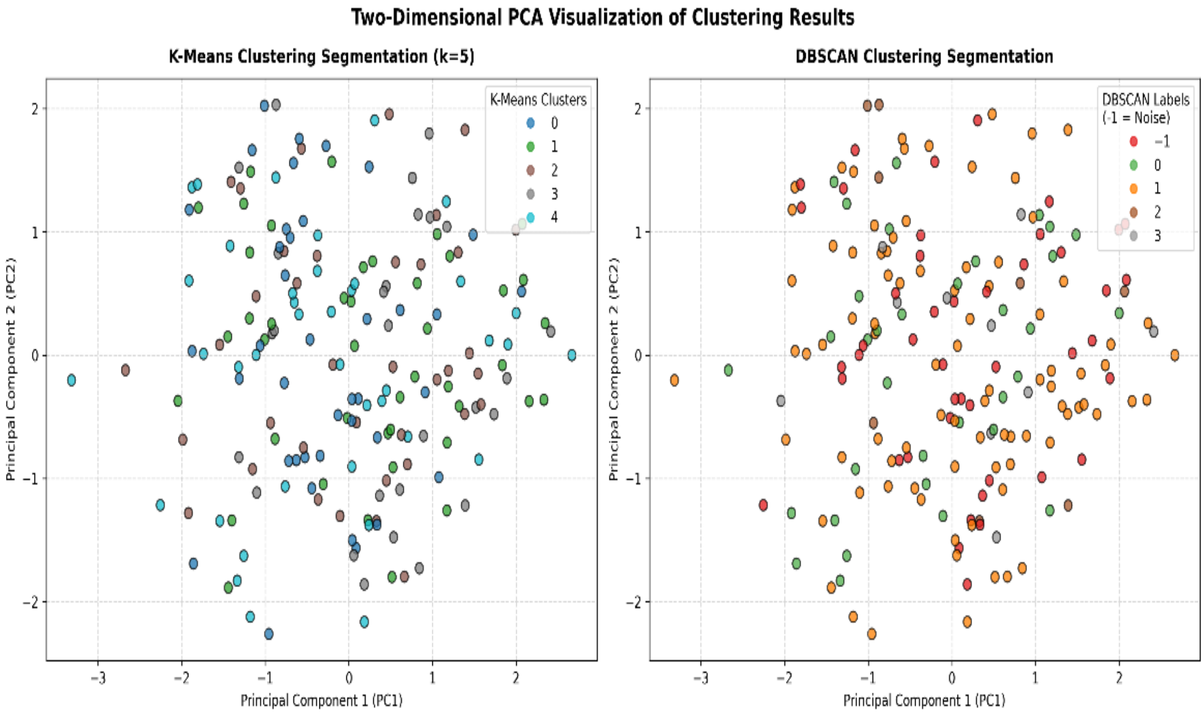
Table 3.
Comparison of Clustering Performance Metrics

Method	Number of Clusters	Noise Points	Silhouette Score	Davies-Bouldin Index
K-Means	5	-	0.3997	0.9999
DBSCAN	4	39	0.4830	0.7240

Source table: Author

Table 3 compares the clustering performance of K-means and DBSCAN. DBSCAN performs better, as indicated by its higher silhouette score (0.4830) and lower Davies-Bouldin Index (0.7240) than K-means (Silhouette = 0.3997; DBI = 0.9999). These results indicate that DBSCAN forms more compact and better–separated clusters while also identifying atypical representatives as noise.

Figure 5.
Two-Dimensional PCA Visualization of K-Means and DBSCAN Clustering Results



Source image: Author

Figure 5 presents a two-dimensional visualization of the clustering results obtained from K-Means and DBSCAN using Principal Component Analysis (PCA). The original five-dimensional feature space was projected into two principal components to facilitate visual inspection of cluster separation. The K-Means results show noticeable overlap among several clusters, indicating that some recipient groups share similar socio-economic and academic characteristics. In contrast, DBSCAN produces denser, more compact groupings while also identifying several observations as noise. These visual findings are consistent with the quantitative evaluation results, where

DBSCAN achieved a higher Silhouette Score (0.483) and a lower Davies–Bouldin Index (0.724) than K-Means (Silhouette = 0.3997; DBI = 0.9999). Therefore, the PCA visualization provides additional empirical support for the conclusion that DBSCAN captures the underlying structure of the recipient data more effectively than K-Means.

Beyond its statistical performance, the DBSCAN results have important ethical and policy implications. Identifying a substantial number of noise observations indicates that some beneficiaries have unique combinations of socioeconomic, academic, and social characteristics that general recipient categories cannot adequately represent. In the context of Islamic social finance, these cases should not be viewed merely as statistical outliers but as individuals who may require special consideration. Therefore, data-driven approaches should serve as complementary decision-support tools rather than substitutes for institutional judgment. Integrating clustering analysis with human assessment can help ensure that efficiency and transparency objectives remain aligned with the principles of fairness, compassion, inclusiveness, and social justice that underlie ZIS distribution.

Several limitations should be considered when interpreting this study's findings. First, the analysis relies on secondary data obtained from Lazismu UAD administrative records. Although these data are suitable for exploratory analysis, they were not originally collected for research purposes, which may limit the availability of potentially relevant socioeconomic indicators such as household income, number of dependents, or detailed family circumstances. Second, the dataset represents recipients from a single higher education institution. Consequently, the resulting clusters may reflect institutional characteristics specific to UAD and may not fully represent ZIS beneficiaries in other universities or Islamic philanthropic organizations. Third, clustering techniques identify patterns of similarity rather than causal relationships. Therefore, interpret the clusters as analytical groupings that support policy evaluation rather than definitive categories for determining eligibility. Finally, institutional procedures and administrative practices may influence the observed allocation patterns. As a result, researchers cannot eliminate potential institutional bias when using secondary data. Future studies are encouraged to incorporate data from multiple institutions, additional socioeconomic variables, and qualitative assessments to provide a more comprehensive understanding of ZIS distribution practices.

Conclusion

This study demonstrates that unsupervised clustering methods can effectively reveal latent structures in the characteristics of zakat, infaq, and sadaqah (ZIS) recipients, providing valuable analytical insights to support data-driven policy decisions. K-Means and DBSCAN show that ZIS beneficiaries are heterogeneous, with variations in socio-economic status, academic performance, and social engagement. While K-Means provides a clear general segmentation, DBSCAN produces more compact, well-separated clusters and identifies atypical cases that require individual policy consideration. Thus, clustering is not meant to replace institutional decision-making but serves as an objective, systematic analytical tool. The findings suggest that integrating clustering-based analysis can enhance transparency, fairness, and accountability in ZIS distribution while remaining sensitive to the diversity of individual recipients.

From a practical perspective, the findings provide several implications for Lazismu UAD. First, Lazismu UAD can incorporate clustering-based recipient segmentation into the beneficiary evaluation process as a decision-support mechanism to complement existing administrative assessments. Second, identifying recipient groups with differing levels of economic vulnerability, academic achievement, and social participation can help prioritize assistance in line with institutional objectives and principles of distributive justice. Third, detecting atypical recipients through DBSCAN highlights the importance of maintaining flexibility in allocation policies so students with unique circumstances are not overlooked. Practically, Lazismu UAD may utilize the clustering results to develop a recipient segmentation dashboard, periodically review allocation patterns, and identify priority groups for further verification before final assistance decisions are made. Consequently, integrating data analytics into ZIS management may strengthen institutional transparency, improve targeting accuracy, and support more evidence-based allocation strategies.

For future research, it is recommended to employ the K-Prototypes partitional clustering method, specifically designed for mixed data containing both numerical and categorical variables. This approach is expected to preserve the semantic meaning of categorical variables, generate more representative clusters, and improve interpretability, thereby providing a stronger basis for formulating adaptive, equitable, and evidence-based ZIS distribution policies. Future studies may also incorporate additional socioeconomic variables and data from multiple institutions to improve the generalizability of clustering results and support broader applications of machine learning in Islamic social finance.

Acknolewgments

The author expresses sincere gratitude to LPPM Universitas Ahmad Dahlan for the financial support that made this research possible.

Author Contributions Statement

Safika Maranti conceptualized the study and led the analysis of ZIS fund distribution based on recipient characteristics. Sugiarto Surono developed and implemented the machine learning models. Khusnul Khotimah contributed optimization techniques and improved model performance. Nurhidayati Kasrul Jalil handled the analysis of social aspects and recipient characteristics. Sintia Afriyani conducted clustering analysis and interpreted the data. All authors contributed to writing, reviewing, and approving the final manuscript.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

AI Usage Statement

The author declares that the use of Artificial Intelligence (AI) in this study was limited to supportive functions, including language editing, grammar checking, and improving clarity and readability. AI was not used to develop research ideas, analyze data, interpret results, or draw conclusions. All intellectual contributions remain the sole responsibility of the author, in accordance with academic ethical standards.

Bibliography

Ahmed, M., Seraj, R., Mohammed, S., & Islam, S. (2020). The k-means Algorithm : A

Comprehensive Survey and Performance Evaluation. *Electronics*, *9*(8), 1–12. https://doi.org/10.3390/electronics9081295

Almaida, N. (2023). Distribusi Dana Zakat, Infak dan Sedekah untuk Penguatan Pendidikan. *Civil Officium : Journal of Empirical Studies on Social Science*, *3*(2), 42–50. https://doi.org/https://doi.org/10.53754/civilofficium Distribusi

Alwi, M., Sarjan, M., & Yusuf, H. (2023). *Digitalisasi Pengelolaan Dana Zakat Dalam Pemberdayaan Ekonomi Umat*. *8*(2), 118–142. https://doi.org/10.35329/jalif.v8i1.3834

Aprianti, W., Permadi, J., Rhomadhona, H., & Amelia, N. (2024). Implementation of KNN and AHP-TOPSIS as a Recommendation System for Mustahik Selection. *Brilliance Research of Artificial Intelligence*, *4*(1), 192–200. https://doi.org/https://doi.org/10.47709/brilliance.v4i1.3883

Aprianto, M. S., Pane, S. F. S., Natasya, T. A., & Handoko, W. (2025). Prediksi Kelayakan Mustahiq Zakat Fitrah di Masjid Nur Hadi Menggunakan Metode Naïve Bayes. *Fusion : Journal of Research in Engineering, Technolgy and Applied Sciences*, *2*(April), 23–30. https://doi.org/https://doi.org/10.66341/fusion.v2i1.276

Baidhowi, M. M., & Triwibowo, A. (2023). Zakat Management Improvement Strategy Efforts to Maintain Muzakki Trust. *Journal of Contemporary Applied Islamic Philanthropy*, *1*(1), 1–10. https://journal.nubaninstitute.org/index.php/jcaip/article/view/9

Bawana, T. A., Abdullah, R., Novriadi, D., S, M. M., Mubarok, A., Susilawati, N., Zulkifli, Haerudin, U. M., Marlina, L., Hadi, R., Hasanuddin, Maranti, S., & Alam, A. S. (2023). *Zakat Era Society 5.0* (Safarinda Imani (ed.)). Media Sains Indonesia.

Chuang, H., & Chou, W. (2025). SilhouetteScoreinR : Beyond traditional network layouts by leveraging local cohesion and nearest neighbor separation. *MethodsX*, *15*(July), 103622. https://doi.org/10.1016/j.mex.2025.103622

Efita, S. D., & Triase. (2024). Optimizing Zakat Distribution with GIS and Data Mining in Community Empowerment at BAZNAS Deli Serdang. *Journal of Information Systems and Informatics*, *6*(4), 3087–3103. https://doi.org/https://doi.org/10.51519/journalisi.v6i4.942

Faiza, N. A. R., Syarifudin, S., & Nurafini, F. (2023). Integrasi Keuangan Sosial (ZISWAF) dan Bisnis Pesantren dalam Pemberdayaan Sosial-Ekonomi Masyarakat di Jawa Timur. *Jurnal Ekonomika Dan Bisnis Islam*, *6*(1), 154–165. https://doi.org/10.26740/JEKOBI.V6N1.P154-165

Fakiriyah, S., Hana, N., Ramadani, A., & Panorama, M. (2023). Analisis Pendistribusian Dana Zakat, Infaq, Dan Sedekah (ZIS) Yang Tidak Produktif Di Indonesia. *Equivalent : Journal Of Economic, Accounting and Management*, *1*(1), 1–8. https://doi.org/https://doi.org/10.61994/equivalent.v1i1.21

Hasan, Y. (2024). Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster. *KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer*, *06*(01), 60–74. https://doi.org/https://doi.org/10.54367/kakifikom.v6i1.3938

Hastuti, Q. A. W. (2014). Urgensi manajemen zakat dan wakaf bagi peningkatan kesejahteraan masyarakat. *Ziswaf*, *1*(2), 379–403.

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (Jima-Ilkom)*, *3*(1), 46–56. https://doi.org/https://doi.org/10.58602/jima-ilkom.v3i1.26

Iskandar. (2019). Zakat sebagai Pengurang Kewajiban Pembayaran Pajak : adilkah bagi umat Islam ? *Jurnal Ius Civile 20*, *3*(1), 20–29. https://doi.org/https://doi.org/10.35308/jic.v3i1.1434

Islami, A. A., Haerani, E., & Nazir, A. (2023). Pengelompokan pembagian zakat dengan menggunakan metode clustering k-means. *Jurnal Computer Science and Information Technology (CoSciTech)*, *4*(1), 154–163. https://doi.org/https://doi.org/10.37859/coscitech.v4i1.4804

Ismail, M. M. Ben, & Alsadhan, N. (2023). Simultaneous Classification and Regression for Zakat. *Applied Sciences*, *13*(5244), 1–15. https://doi.org/https://doi.org/10.3390/app13095244

Mohammad, S. A. N., & Maulida, F. E. (2020). Pendayagunaan Dana ZIS Pada Program Mahasiswa Cerdas Baznas (BAZIS) Provinsi DKI Jakarta. *Al-Mizan: Jurnal Hukum Dan Ekonomi Islam*, *4*(1), 54–73. https://doi.org/https://doi.org/10.33511/almizan.v4n1.54-73

Monko, G. J., & Kimura, M. (2023). Optimized DBSCAN Parameter Selection : Stratified Sampling for Epsilon and GridSearch for Minimum Samples. *Software Engineering And Automation*, 43–61. https://doi.org/10.5121/csit.2023.132004

Nabila, N. (2022). Metode fundraising dalam peningkatan dana Zakat Infak Sedekah (ZIS) pada Lembaga Amil Zakat Nahwa Nur Kabupaten Bogor Provinsi Jawa Barat. In *Repository.Uinjkt.Ac.Id*.

Nurrahman, A. N., Nugroho, E. P., & Hambali, Y. A. (2025). Sistem Pendukung Keputusan Penerima Bantuan Zakat Menggunakan Random Forest dan Fuzzy Analytical Hierarchy Process. *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, *4*(2). https://doi.org/https://doi.org/10.62712/juktisi.v4i2.510

Pratama, S. D. (2023). The Role of Zakat in Alleviating Multidimensional Poverty. *International Journal of Islamic Economics and Finance (IJIEF)*, *6*(1). https://doi.org/10.18196/ijief.v6i1.17006

Rahmadi, J., & Mu'minin, A. M. (2024). Sistem Pendukung Keputusan Penyaluran Zakat dengan Metode Simple Additive Weighting (SAW) di LAZ Musa ' adatul Ummah Bandung. *SisInfi : Jurnal Sistem Informasi Dan Informatika*, *6*(2), 66–75. https://doi.org/https://doi.org/10.37278/sisinfo.v6i2.913

Rahmawati, M., & Yazid, M. (2025). Implementasi Ziswaf Dalam Pengembangan Ekonomi Umat Melalui Maqashid Syariah. *Jurnal Ilmiah Mahasiswa Ekonomi Syariah (JIMESHA)*, *5*(1), 41–52. https://doi.org/https://doi.org/10.36908/jimesha.v5i1.583

Rao, K. M., Saikrishna, G., & Supriya, K. (2023). Data preprocessing techniques: emergence and selection towards machine learning models - a practical review using HPA dataset. *Multimedia Tools and Applications*, *82*(24), 37177–37196. https://doi.org/https://doi.org/10.1007/s11042-023-15087-5

Saiful, M., Nur, A. M., Sutriandi, A. E., Nuzululnisa, B. N., & Puspita, E. (2025). Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Klasifikasi Penyaluran Dana Zakat Muhammad. *Infotek: Jurnal Informatika Dan Teknologi*, *8*(1), 239–249. https://doi.org/https://dx.doi.org/10.29408/jit.v8i1.28624

Saputri, E. (2025). Teknik dan aplikasi data mining di Indonesia : tinjauan literatur satu dekade (2015-2024). *IT-EXPLORE Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, *04*(2), 138–149. https://doi.org/10.24246/itexplore.v4i2.2025.pp138-149

Simbolon, I. N., & Friskila, P. D. (2024). Analisis dan Evaluasi Algoritma DBSCAN (Density-Based Spatial Clustering Of Applications With Noise) Pada Tuberkulosis. *JITET (Jurnal Informatika Dan Teknik Elektro Terapan)*, *12*(3), 3778–3787. https://doi.org/http://dx.doi.org/10.23960/jitet.v12i3S1.5206

Taha, K., Yoo, P. D., Yeun, C., Homouz, D., & Taha, A. (2024). A comprehensive survey of text classification techniques and their research applications : Observational and experimental insights. *Computer Science Review*, *54*(July), 100664. https://doi.org/10.1016/j.cosrev.2024.100664

Triayudi, A. (2023). Penerapan Data Mining Untuk Klasifikasi Penerima Dana Bantuan Sosial Dengan Menggunakan Algoritma K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, *5*(2), 532–542. https://doi.org/10.47065/bits.v5i2.3972