Journal of Applied Engineering and Technological Science
Vol 6(1) 2024: 192-205

INTERACTIVE GEOGRAPHIC VISUALIZATION AND UNSUPERVISED LEARNING FOR OPTIMAL ASSIGNMENT OF PREACHERS TO APPROPRIATE CONGREGATIONS

Rahmad Kurniawan[1*], Ibnu Daqiqil Id[2], Abdul Somad Batubara[3], Fitra Lestari[4], Arisman Adnan[5], Fatayat[6], Ilyas Husti[7]
Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Riau[1,2,6]
Universiti Islam Sultan Sharif Ali, Bandar Seri Begawan, Brunei Darussalam[3]
Department of Industrial Engineering, Universitas Islam Negeri Sultan Syarif Kasim Riau[4]
Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Riau[5]
Graduate Program, Universitas Islam Negeri Sultan Syarif Kasim Riau[7]
rahmadkurniawan@lecturer.unri.ac.id[1*]

Received: 28 July 2024, Revised: 27 October 2024, Accepted: 02 November 2024
**Corresponding Author*

ABSTRACT
Riau Province has a population of 6,642,874 and a diverse geography, which poses significant challenges in optimizing Islamic preaching activities. Traditional assignment methods often lead to inefficiencies due to misalignment between the preacher's expertise and congregational needs, as well as logistical issues. This study integrates K-Means clustering and DBSCAN algorithms with interactive geographic visualization to optimize the assignment of preachers to mosques. We collected 435 data points, including 185 mosques and 250 preachers. K-Means was evaluated using the Elbow Method and Silhouette Score, identifying 10 clusters as optimal with a Silhouette Score of 0.435654. However, K-Means does not handle outliers effectively, as indicated by zero outliers in all configurations. DBSCAN was tested with various epsilon (eps) and minimum sample values. The optimal configuration with eps of 1.5 and 5 minimum samples resulted in 10 clusters with a Silhouette Score of 0.381108 and 60 outliers. DBSCAN effectively manages outliers and varying densities. Although K-Means is advantageous for its simplicity and higher Silhouette Scores, it is unable to handle outliers effectively. DBSCAN provides robust clustering for noisy data. Therefore, it can be concluded that hybridizing unsupervised learning algorithms with geographic visualization can potentially improve the effectiveness of preaching activities in Riau Province and enhance preacher assignment.
Keywords: *Preacher Assignment, Optimization, K-Means Clustering, DBSCAN, MUI Riau.*

1. Introduction

The assignment of preachers to various congregations for delivering *da'wah* (Islamic preaching) is a critical function of the Indonesian Ulema Council (MUI) Riau Province Chapter. Effective assignment ensures that Islamic preaching is disseminated in a relevant, timely manner and aligned with the community's needs. In Riau Province, with a population of 6,642,874 (Province, 2024) and its diverse and varied geographic landscape, the challenge of optimal preacher deployment is particularly pronounced. Traditional methods of assigning preachers often result in inefficiencies, such as misalignment between the preacher's expertise and the congregation's needs, as well as logistical challenges related to travel and accessibility (Nur et al., 2021).

By contrast, the application of machine learning and data analytics in various fields has revolutionized traditional practices by providing data-driven insights and optimization strategies. Unsupervised learning algorithms, such as K-means and density-based spatial clustering of applications with noise (DBSCAN), have shown great promise in clustering and analyzing complex datasets (Hu et al., 2023; Raja et al., 2024). These techniques can uncover hidden patterns and relationships within the data, leading to more informed decision-making processes (Kurniawan, Abdullah, Lestari, Nazri, Mujahidin, et al., 2020; Nur et al., 2021). Leveraging these technologies for the assignment of preachers could significantly enhance the effectiveness and efficiency of preaching activities.

Business intelligence has also emerged as a powerful tool for visualizing spatial data and facilitating interactive analysis (Hengki et al., 2021; Lavalle et al., 2019). Business intelligence also enhances decision-making, improves business processes, reduces costs, understands customer needs, and provides a competitive advantage (Fernandes et al., 2023; Tang et al., 2022). Integrating business intelligence with unsupervised learning algorithms allows for the creation of interactive maps that can visually represent the distribution of preachers and mosques. Such visualizations can support identifying optimal locations for preacher assignments, considering factors such as geographic proximity, accessibility, and the specific needs of different congregations. The dataset includes geographic coordinates, such as latitude and longitude, and operational features of mosques, including their accessibility and the types of religious activities they hold. Applying K-means and DBSCAN algorithms to this dataset generates clusters representing groups of mosques with similar characteristics. These clusters are then visualized on interactive maps using *Folium*, a powerful Python library for creating leaflet maps. This visualization provides an intuitive way to explore the data and supports decision-making in preacher assignments.

To optimize the assignment of preachers in Riau Province, we provide a novel approach that combines interactive geographic visualization and unsupervised learning algorithms in this work. We aim to construct a more efficient and effective system for preacher assignments by clustering mosque and preacher data based on geographic and accessibility features and visualizing these clusters on interactive maps. This approach not only addresses the logistical challenges but also ensures that the religious teachings are more relevant to the specific issues faced by different communities. Based on feature similarity, the K-means algorithm is a well-liked clustering technique that divides data into an established number of clusters. It has been widely used in various fields for its simplicity and effectiveness in uncovering patterns in large datasets. However, its performance can be influenced by the initial choice of cluster centers and the assumption that clusters are spherical. To complement K-means, we also employ DBSCAN, which identifies clusters based on density and can handle arbitrary-shaped clusters and noise, making it suitable for more complex datasets as in the case study of religion.

In recent years, research on clustering algorithms, particularly the K-means and DBSCAN algorithms has seen significant development and application across various domains. The K-means algorithm has been widely recognized as a powerful tool in data mining (Ahmed et al., 2020; Hu et al., 2023). K-means is known for its effectiveness in solving complex problems such as religion (Nur et al., 2021), health (Kurniawan, Abdullah, Lestari, Nazri, Akhmad, et al., 2020), urban road planning (Ran et al., 2021), student GPA prediction (Santosa et al., 2021), organizational data classification, and movie recommendation systems (Cintia Ganesha Putri et al., 2020). The DBSCAN algorithm, a density-based clustering method, has gained attention for its ability to handle outliers and cluster non-convex data (Balavand, 2022). Studies have focused on optimizing DBSCAN through arithmetic optimization algorithms (Yang et al., 2022) and proposing new variations like NS-DBSCAN for network space clustering (T. Wang et al., 2019). Additionally, a study has explored parameter determination methods for DBSCAN, such as using improved metaheuristic algorithms (Lai et al., 2019).

Even though one of the algorithms ultimately proved to be the better algorithm for our specific dataset and objectives however several proper reasons exist for including experiments with K-means and DBSCAN in our research. This comparison demonstrates the specific advantages of an algorithm, particularly in handling outliers and varying densities. In real-world applications, data characteristics are not always known in advance, and exploring multiple algorithms ensures that the chosen method is robust and well-suited to the specific data characteristics. Therefore, this study employs the K-means and DBSCAN algorithms to leverage their unique strengths and ensure the best clustering model. K-means is favored for its simplicity, efficiency, and applicability in various fields. However, it requires careful parameter selection and is sensitive to outliers (Ran et al., 2021). DBSCAN is preferred for its ability to handle noise and identify clusters of varying shapes (Balavand, 2022). By utilizing these algorithms, this study can benefit from the strengths of both methods, achieving more robust and accurate clustering results across different datasets and applications. This study leverages the strengths of both algorithms to address a wide range of clustering challenges in this domain.

2. Literature Review

In recent years, the application of machine learning in general and unsupervised learning in particular has seen significant progress in various fields. Clustering algorithms such as K-means and DBSCAN have been widely used in various fields because they cluster unlabeled data into groups with similar characteristics. In the field of Islamic case studies, this algorithm has been used more frequently to optimize services to the community. In Islamic studies and applications, the latest studies have witnessed notable advancements in research utilizing clustering algorithms such as K-means and DBSCAN. These algorithms have been applied in various Islamic fields, including Islamic finance, Islamic banking, and Islamic case studies, to extract meaningful insights and patterns. Researchers have explored the integration of the K-means algorithm with genetic algorithms to enhance clustering quality (Islam et al., 2018). Moreover, the K-means algorithm has been utilized in predicting determination clusters, demonstrating its versatility in diverse applications within Islamic studies (Wahyuni et al., 2023).

The DBSCAN algorithm has been employed in Islamic finance to analyze product utilization by Islamic retail banks (Otaviya & Rani, 2020). Its ability to handle outliers and identify clusters of varying densities has been leveraged to gain insights into the productivity and determinants of Islamic banks in Indonesia (Otaviya & Rani, 2020). DBSCAN has also been used in Islamic case studies to investigate the shape of structures formed by self-assembly processes, showcasing its applicability beyond traditional clustering domains (HabiBoğlu et al., 2022). The K-means and DBSCAN algorithms have been particularly beneficial in Islamic finance research, where the algorithms have been employed to analyze outlier behavior in customer behavior models (Monalisa & Kurnia, 2019). The studies have extracted valuable insights from complex datasets in Islamic contexts by leveraging both algorithm's strengths and contributing to the advancement of knowledge in Islamic studies and applications. The latest studies have seen a surge in research utilizing clustering algorithms like K-means and DBSCAN in Islamic fields and demonstrating their effectiveness in extracting meaningful patterns and insights from data in various domains within Islamic studies.

The use of clustering algorithms such as K-means and DBSCAN in various fields is increasing. However, there is still a research gap in the specific application of these algorithms, namely for the optimal assignment of preachers in mosques that are close in terms of location, according to their expertise and easy access. Although existing literature has explored the use of clustering algorithms in diverse domains such as medical image segmentation and risk-based *sharia* audit implementation (Sani & Abubakar, 2021), studies focusing on the fitted application of K-Means and DBSCAN algorithms in the context of Islamic preaching are lacking. The current body of research predominantly emphasizes the technical aspects and general applications of clustering algorithms, overlooking the unique requirements and considerations involved in assigning preachers to congregations within Islamic educational settings. By bridging this research gap, we develop a specialized framework that integrates the K-means and DBSCAN algorithms to optimize the assignment of preachers based on factors specific, such as geographic proximity, expertise in Islamic jurisprudence, accessibility, and compatibility with congregational needs.

Furthermore, recent studies have shown a growing trend in combining machine learning with data visualization to enhance various fields. Researchers have recognized the potential of machine learning in automating visualization design and analysis (Dibia & Demiralp, 2019). This integration aims to help individuals understand, diagnose, improve, and apply machine learning models more effectively (Q. Wang et al., 2019). Visual analytics, which combines machine learning techniques with visualizations and interaction, has been utilized to explore data and assist in analytical tasks (Q. Wang et al., 2019).

We propose to improve the overall effectiveness of Islamic preaching activities in Riau Province by adopting this innovative approach. The optimized assignment of preachers not only ensures that religious teachings are more relevant and impactful but also enhances the efficiency of preachers deployment, reducing travel time and logistical challenges. This study provides a framework that can be adapted and implemented in other regions and contexts, contributing to the broader goal of leveraging technology to support community and religious services.

Interactive geographic visualization tools, such as *Folium* have become increasingly popular because of the ability to create dynamic maps that enhance data exploration and decision-making. Applying such tools in the context of religious services can provide stakeholders with a clear and comprehensive view of preacher's distribution and mosque locations, facilitating more informed and effective decision-making. By integrating these tools with clustering algorithms, this study aims to bridge the gap in the existing literature and provide a robust framework for optimizing preacher assignments.

3. Research Methods

We intend to improve the overall effectiveness of preaching activities in Riau Province by adopting a comprehensive methodology. Optimizing preacher assignments ensures that preaching is more relevant and impactful and enhances the efficiency of preacher deployment, reducing travel time and logistical challenges. This study provides a framework that is expected to be implemented in other regions. Thus, it is expected to contribute widely to technology that can support services to the Muslim community.

Therefore, this study employs a systematic methodology to optimize the assignment of preachers to mosques using unsupervised learning algorithms and geographic visualization tools. The stages of this study involved several steps such as data collection and data preprocessing, clustering modelling using K-Means and DBSCAN, and evaluation of clustering results with suitable parameters. The last step was cluster visualization using an interactive map. Figure 1 below explains the brief of the research methodology. Each step is designed to ensure that the data is analyzed and presented in a way that supports efficient and effective preacher's assignments.

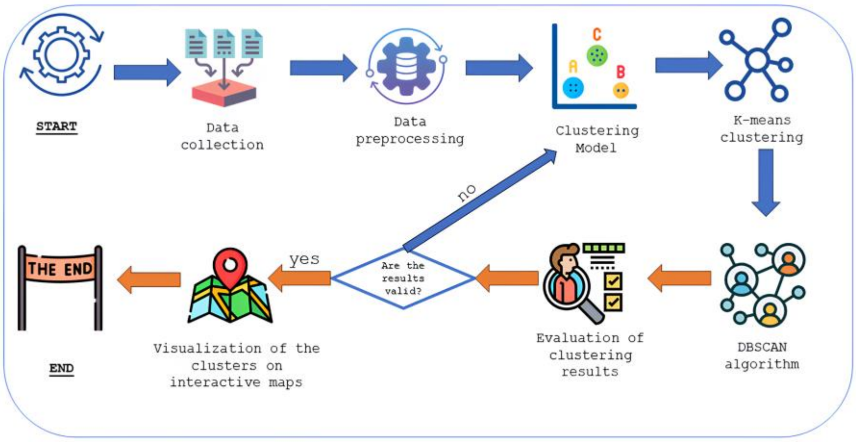

Fig. 1. Methodology for Interactive Geographic Visualization and Unsupervised Learning for the Optimal Assignment of Preachers to Congregations

Data Collection and Preprocessing

The data used in this study was obtained from the Indonesian Ulema Council (MUI) Riau Province Chapter and includes detailed information on preachers and mosques in Riau Province. We collected 435 data points, including 185 mosques and 250 preachers. The dataset comprises geographic coordinates (latitude and longitude), accessibility features (such as proximity to major roads and the presence of narrow alleys), and various operational characteristics of the mosques (such as the frequency of different types of religious activities).

Let $X = \{x_1, x_2, \ldots, x_n\}$represent the dataset, where each x_i is a vector containing the features of mosques. Initially, if the dataset contained missing values, they were replaced with the mean of the non-missing values in the respective feature column. Specifically, for each missing value f_i in a feature column *f* of *X*, each missing value f_i was replaced as follows, using Equation 1:

$$f_i = \begin{cases} \frac{1}{n}\sum_{j=1}^{n} f_j, & \text{if } f_i \text{ is } NaN \\ f_i, & otherwise \end{cases} \qquad (1)$$

Additionally, to prepare the data for clustering, all features were normalized using the *StandardScaler* from the *scikit-learn library*, ensuring that each feature contributes equally to the clustering process. Normalize the feature vectors using *StandardScaler* as written in Equation 2, where μ is the mean of *X* and σ is its standard deviation:

$$scaled_x = \frac{\mathrm{X} - \mu}{\sigma} \qquad (2)$$

Clustering Model using K-Means Clustering and DBSCAN Algorithm

We applied two clustering algorithms, i.e., K-Means and DBSCAN to identify groups of preachers and mosques with similar characteristics. The K-Means algorithm partitions the data into a predefined number of clusters by minimizing the within-cluster variance. To determine the optimal number of clusters (k), we used the Elbow Method and Silhouette Score, which evaluate cluster cohesion and separation. This method involved computing the Sum of Squared Errors (*SSE*) for each (k). The *SSE* is calculated using the Equation 3:

$$SSE(k) = \sum_{i=1}^{n} min_{\mu j \in C} \mid\mid x_i - \mu_j \mid\mid^2 \qquad (3)$$

where x_i represents each data point (each mosque and preacher), and μj represents the centroid of cluster j. The optimal (k) is identified at the point where the rate of decrease in SSE slows down significantly, indicating diminishing returns from adding more clusters. Next, we computed the Silhouette Score for each (k) to evaluate the quality of the clustering. The Silhouette Score is given by:

$$Silhouette(k) = \frac{1}{n}\sum_{i=1}^{n} \frac{b_i - a_i}{\max(a_i, b_i)} \qquad (4)$$

where a_i is the average distance between x_i and all other points in the same cluster, and b_i is the average distance between x_i and all points in the nearest cluster. This metric helps in understanding how well each data point lies within its cluster. Finally, the clustering assignment is made by assigning each mosque and preacher to one of the (k) clusters:

$$Cj = \{x_i \colon \mid\mid x_i - \mu_j \mid\mid \leq \mid\mid x_i - \mu_j \mid\mid \forall l \neq j\} \qquad (5)$$

This method ensures that each preacher is assigned to the nearest cluster based on the computed centroids, thereby optimizing the grouping.

Meanwhile the DBSCAN algorithm does not need the number of clusters to be specified a priori to identify clusters based on density. This method allows it to find clusters of random shapes and effectively handle noise. We experimented with diverse values of the epsilon (ϵ) parameter and the minimum number of samples to determine the best configuration for our data For DBSCAN clustering, parameter selection is crucial. Different values for $\epsilon\ (eps)$ and $MinPtsMinPts\ (\min\ samples)$ were explored to determine the optimal parameters. The optimal parameters are those that maximize the Silhouette Score:

$$(\epsilon, MinPts) = \arg max_{\epsilon, MinPts}\, Silhouette(\epsilon, MinPts) \qquad (6)$$

Once the optimal parameters were selected, clusters were formed. For each point x_i, the neighborhood $N_\epsilon(xi)$ was identified using the Equation:

$$N_\epsilon(x_i) = \{x_j \in X \colon \mid\mid x_i | - x_j \mid\mid \leq \epsilon\} \qquad (7)$$

A point x_i is classified as a core point if:

$$|N_\epsilon(x_i)| \geq MinPts \qquad (8)$$

as a border point if:

$$1 < |N_{\epsilon}(x_i)| < MinPts \qquad (9)$$

or as noise if:

$$|N_{\epsilon}(x_i)| = 1 \qquad (10)$$

This study process was carried out to ensure that clusters with varying densities were correctly identified, including handling noise and outliers that were resolved effectively. By following this process, this study can optimize the assignment of preachers in mosques in Riau Province effectively and efficiently.

Parameters Tested

The parameters tested for K-Means and DBSCAN clustering are summarized in Table 1 below:

Table 1 - K-Means and DBSCAN Parameters Tested

Algorithm	Parameter	Values Tested
K-Means	Number of Clusters	1-10
DBSCAN	eps	0.1, 0.5, 1.0, 1.5, 2.0
DBSCAN	Min Samples	5, 10, 15, 20

We can effectively group mosques based on geographic data and profile similarities by systematically applying this clustering algorithm, thus facilitating better assignment of preachers according to the congregation's needs. This structured approach increases the impact and reach of the overall Islamic preaching activities.

Evaluation of Clustering Results

The performance of the clustering algorithms was evaluated using several metrics. For K-Means, we used the Elbow technique to analyze the Sum of Squared Errors (SSE) and determine the point where adding more clusters did not significantly improve the model. The SSE measures the within-cluster variance. Lower SSE values indicate the best clustering. In Equation 3, x_i represents each data point (each mosque and preacher), and μj represents the centroid of cluster j.

We also used the Silhouette Score to measure how similar each point in a cluster is to the points in its cluster assessed to the points in other clusters. The Silhouette Score is calculated as Equation 4, where a_i is the average distance between x_i and all other points in the same cluster, and b_i is the average distance between $x_i$and all points in the nearest cluster. Higher Silhouette Scores indicate better clustering performance, as they reflect greater cluster cohesion and separation.

In DBSCAN, the Silhouette Score was calculated for various combinations of eps and $min\_samples$ values to identify the optimal settings. The optimal parameters are determined by maximizing the Silhouette Score. Additionally, the number of clusters and the number of outliers detected were recorded. DBSCAN's ability to handle noise is evaluated by counting the number of points classified as noise. The neighborhood for each point x_i is identified using Equation 7-10. Table 2 shows a comprehensive evaluation of the clustering algorithms and ensuring that the clustering results are both effective and meaningful.

Table 2 - Clustering Evaluation Metrics

Algorithm	Metric	Description
K-Means	SSE	Measures within-cluster variance.
K-Means	Silhouette Score	Measures cluster cohesion and separation. Higher values indicate better clustering
DBSCAN	Silhouette Score	Similar to K-Means, but for density-based clustering
DBSCAN	Number of Clusters	Number of distinct clusters identified
DBSCAN	Number of Outliers	Number of points classified as noise

Visualization of Clustering Results

We used *Folium*, a powerful Python library for creating interactive maps to visualize the clustering results. The clusters were plotted on a geographic map, each represented by a diverse color. Markers were added for each preacher and mosque, displaying the name and cluster number upon clicking. This interactive map provides an intuitive way to explore the data and supports decision-making in the assignment of preachers. The use of interactive maps based on Business Intelligence can also increase transparency and allow for ongoing real-time data updates and adjustments based on community needs.

The visualizations were designed to highlight the spatial distribution of preachers and mosques and the clustering results, making it easier to identify patterns and areas that require attention. This approach facilitates a more effective and informed deployment of preachers, ensuring that their expertise matches the specific needs of different congregations.

4. Results and Discussions

In the K-means algorithm, we use appropriate validation techniques, such as Elbow and Silhouette Scores to determine the optimal number of clusters. These methods allowed us to evaluate the cluster cohesion and separation to find the most effective clustering solution. The Elbow technique contains plotting the Sum of Squared Errors (SSE) against the number of clusters. The point where the SSE decreases slower, forming an elbow suggests the optimal number of clusters. The Silhouette Score can provide information about how well each data point fits into a given cluster. A higher score indicates that a more precise cluster is formed. Table 3 presents the results for the Silhouette Score, SSE, and the number of outliers across different cluster numbers.

Table 3 - K-Means Results for Silhouette Score, SSE, and the Number of Outliers Across Different Cluster Numbers

Number of Clusters	Silhouette Score	SSE	Number of Outliers
1	NaN	2212	0
2	0.375493	1578.39	0
3	0.298832	1300.41	0
4	0.308033	1122.607	0
5	0.334958	963.7137	0
6	0.334172	857.8206	0
7	0.37055	739.1593	0
8	0.384242	667.702	0
9	0.409541	618.3339	0
10	0.435654	554.2245	0

Based on Table 1, the SSE values decrease as the number of clusters increases, which is expected because adding more clusters can reduce the within-cluster variance. However, the rate of decrease in SSE becomes less significant beyond a certain point. Furthermore, Figure 2 shows that the elbow point indicates the optimal number of clusters. In this case, the elbow appears to be around 4 clusters.

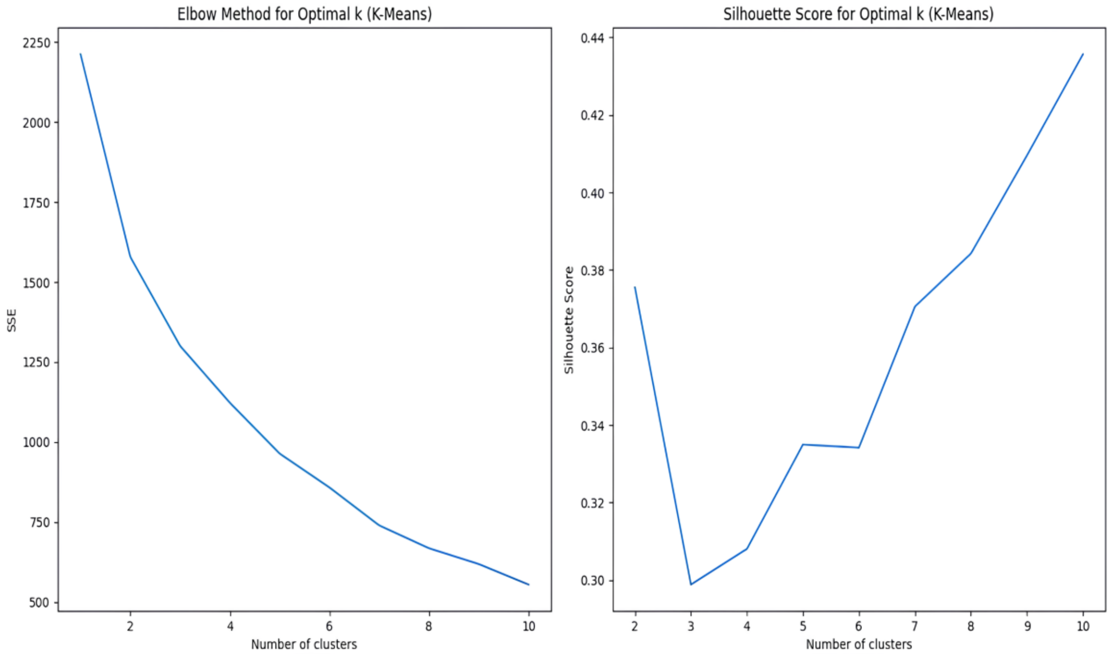

Fig. 2. The Elbow Method and Silhouette Score for Optimal *k* in the K-Means Algorithm

The Silhouette Score provides a complementary perspective. Higher Silhouette Scores indicate that the clusters are well-separated and cohesive. The table shows that the Silhouette Score peaks at 10 clusters with a value of 0.435654. These results suggest that 10 clusters provide the best-defined clusters regarding cohesion and separation, as shown in Figure 3. However, practical considerations such as interpretability and the specific application context should also be considered when determining the final number of clusters.

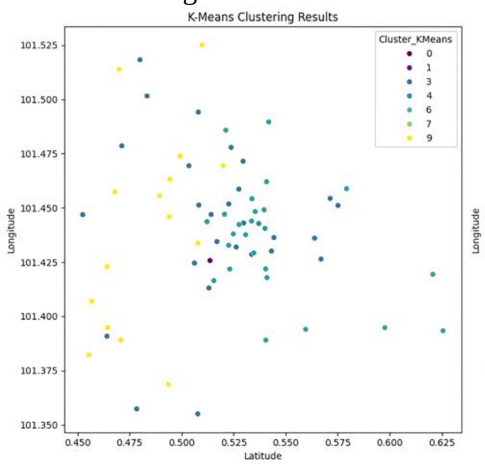

Fig. 3. Visualization of K-Means Clustering Results with 10 Clusters

The K-Means is sensitive to outliers and does not effectively handle them. This result is reflected in Table 3, where the number of outliers is consistently zero across all cluster configurations. These results indicate that the algorithm does not account for noise or outliers in the data, which can affect the clustering performance.

While the Elbow Method suggests that 4 clusters might be optimal based on the SSE, the highest Silhouette Score is achieved with 10 clusters, as shown in Figure 2. Selecting the highest number of clusters as indicated by the Silhouette Score of 10 clusters is considered a practical choice. This approach ensures a balanced distribution of preachers and mosques, aligning with the objective of optimizing assignments while maintaining fairness in the clustering process. Therefore, 10 clusters represent the most optimal solution for our dataset in this context as they balance both within-cluster variance and between-cluster separation. By identifying the optimal number of clusters, we ensure that the clustering results are compelling, meaningful, and aligned with practical application goals.

We subsequently modeled the data using the DBSCAN algorithm. DBSCAN is particularly effective for datasets with noise and varying densities, making it suitable for our clustering needs. Table 4 presents the results of the DBSCAN clustering with various combinations of the epsilon (*eps*) parameter and the minimum number of samples (*min_samples*).

Table 4 - DBSCAN Clustering Results

EPS	Min Samples	Number of Clusters	Silhouette Score	Number of Outliers
0.1	5	7	0.280622	91
0.1	10	2	0.136151	120
0.1	15	1	NaN	130
0.1	20	1	NaN	130
0.5	5	8	0.319766	79
0.5	10	2	0.136151	120
0.5	15	1	NaN	130
0.5	20	1	NaN	130
1	5	9	0.360647	65
1	10	3	0.172466	103
1	15	1	NaN	130
1	20	1	NaN	130
1.5	5	10	0.381108	60
1.5	10	3	0.187389	100
1.5	15	2	0.124092	110
1.5	20	1	NaN	130
2	5	10	0.379065	59
2	10	3	0.189806	99
2	15	2	0.124092	110
2	20	1	NaN	130

The results indicate that the performance of the DBSCAN algorithm varies significantly based on the chosen *epsilon* and *min_sample* values. The number of clusters identified by DBSCAN ranges from 1 to 10. Lower epsilon values (e.g., 0.1) tend to result in fewer clusters, especially with higher *min_samples* values, as shown in Figure 4. Conversely, higher epsilon values (e.g., 1.5 and 2) generally result in a higher number of clusters, demonstrating DBSCAN's sensitivity to the density of the data. For instance, with an epsilon of 1.5 and Min Samples of 5, DBSCAN identified 10 clusters, whereas an epsilon of 0.1 with the same Min Samples identified only 7 clusters.

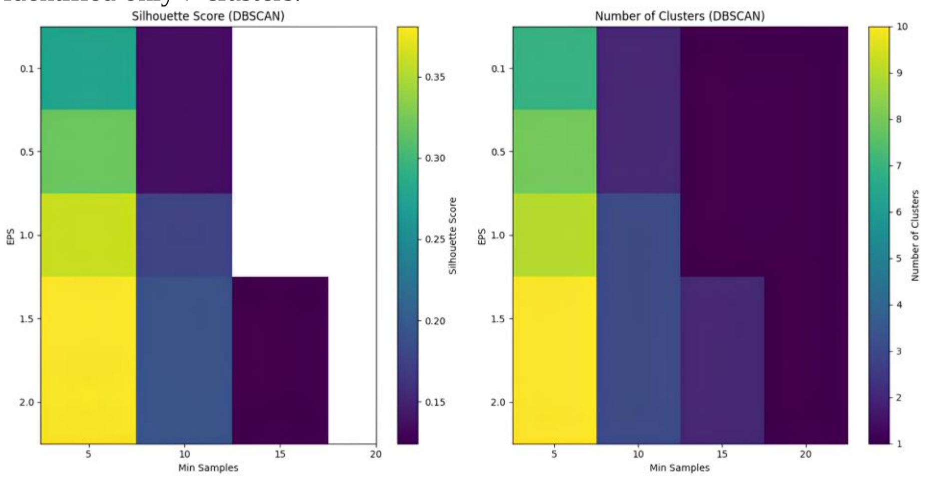

Fig. 4. Visualization of DBSCAN Clustering with Various Combinations of the Epsilon (*eps*) Parameter and the Minimum Number of Samples

The Silhouette Score, which measures the quality of the clustering also varies with different parameter settings. The highest Silhouette Score observed is 0.381108 achieved with an epsilon of 1.5 and Min Samples of 5, indicating well-defined clusters as visualized in Figure 5. Most configurations show moderate clustering quality, with higher Silhouette Scores suggesting better-defined clusters. Configurations resulting in only one cluster have undefined (*NaN*) Silhouette Scores since the measure is not meaningful for a single cluster.

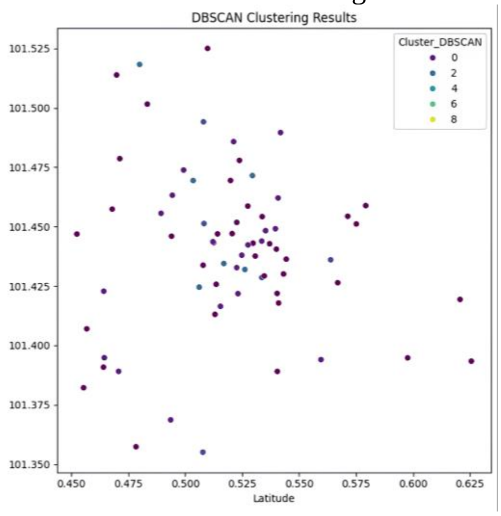

Fig. 5. Visualization of DBSCAN Clustering Results with 10 Clusters

The number of outliers detected by DBSCAN varies considerably. Lower epsilon values and higher *min _sample* values result in more outliers. The maximum number of outliers detected is 130, occurring in configurations that produced only a single cluster indicating that many points were considered noise by the algorithm. For instance, an epsilon of 0.1 with *min _sample* of 20 resulted in 130 outliers, while an epsilon of 2 with *min _sample* of 5 resulted in 59 outliers.

Based on the Silhouette Scores, the configuration with an epsilon of 1.5 and min _sample of 5 produced the best-defined clusters, achieving a Silhouette Score of 0.381108 and identifying 10 clusters with 60 outliers. These results suggest that DBSCAN can effectively identify meaningful clusters and handle noise in the data when the parameters are appropriately tuned.

Through this DBSCAN algorithm, we have considered outlier data and varying densities, thus producing more robust clustering. This result enhances the assignment of preachers to mosques, ensuring that the clusters are well-defined and relevant to the specific needs of different congregations in Riau Province.

Comparative Analysis of K-Means and DBSCAN Clustering Results

A comparative analysis of the results obtained from the K-Means and DBSCAN algorithms is carried out to determine the most effective clustering algorithm for placing preachers in mosques in Riau Province. This section discusses in detail the performance of each algorithm based on the Silhouette Score, the number of clusters formed, and the handling of outlier data. The K-Means algorithm is also evaluated using the Elbow Method and Silhouette Score to identify the optimal number of clusters. The optimal number of clusters determined by the highest Silhouette Score was identified as 10, with a score of 0.435654. However, K-Means does not inherently handle outliers, as indicated by the consistent number of outliers being zero across all configurations. This limitation could compromise clustering quality in the presence of noise within the dataset.

Meanwhile, the DBSCAN algorithm was evaluated using various combinations of the epsilon (eps) parameter and the minimum number of samples (Min Samples). The optimal configuration for DBSCAN was found with an epsilon of 1.5 and Min Samples of 5, resulting in 10 clusters with a Silhouette Score of 0.381108 and 60 outliers. The variability in the number of outliers across different configurations highlights DBSCAN's capacity to manage outliers and adapt to varying densities.

Both K-Means and DBSCAN exhibit distinct strengths and limitations. K-Means is advantageous for its simplicity and higher Silhouette Scores, indicating well-defined clusters. However, it cannot handle outliers effectively. On the other hand, DBSCAN excels in managing outliers and varying densities, providing a more flexible clustering solution for noisy data.

DBSCAN proves to be a more suitable algorithm for datasets with outliers, owing to its robustness in handling outliers and detecting clusters with varying shapes and densities. This finding ensures a more accurate and practical assignment of Islamic scholars to mosques, enhancing the effectiveness of da'wah activities in Riau Province.

Visualization Results Using Interactive Maps

We successfully applied the optimal clustering models using the K-Means and DBSCAN algorithms to cluster the data on mosques and preachers. The interactive map visualization, shown in Figure 6, illustrates the results of this clustering process.

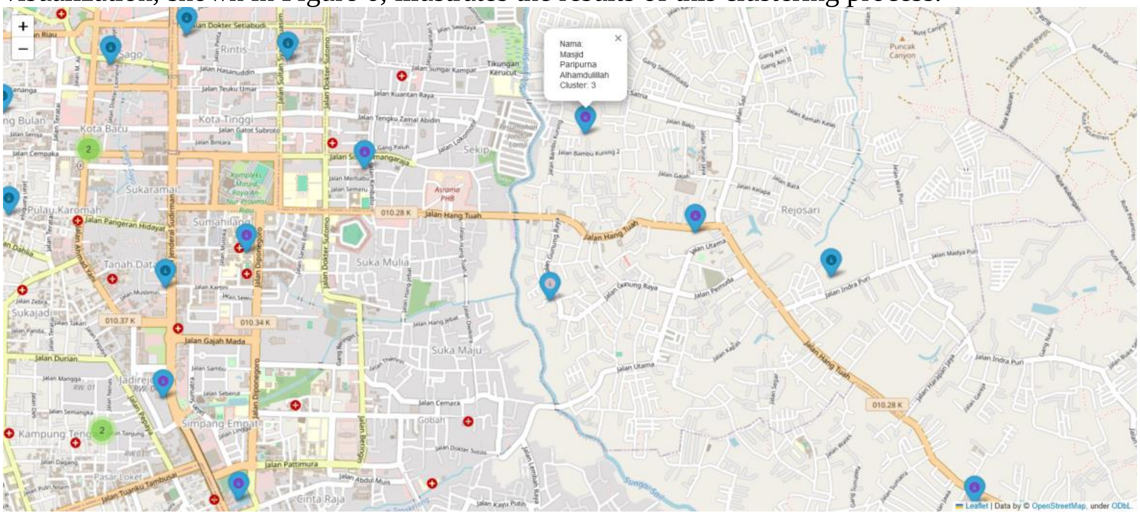

Fig. 6. Example of Clustering Results Using Interactive Maps Visualization

The interactive map displays the mosque's geographical locations and the scholars' residences using pins. When clicked, these pins reveal detailed information about the corresponding mosque or scholar. This feature provides comprehensive data to the mosque management and the authorities at the Indonesian Ulama Council Riau Province Chapter, facilitating better decision-making regarding the assignment of scholars.

The visualization lets users quickly identify which mosques and scholars belong to the same cluster. This clustering information is crucial for recommending assignments, as scholars and mosques within the same cluster are presumed to share similar characteristics or profiles. Therefore, the system can recommend that scholars be assigned to mosques within the same cluster, ensuring alignment between the mosques' needs and the preacher's expertise.

By leveraging the interactive map, the management can visually assess the proximity and cluster affiliation of scholars and mosques. This geographical and operational insight enhances the efficiency and effectiveness of preacher assignments, ensuring that religious teachings are relevant and accessible to the community. The use of this technology not only optimizes the deployment of preachers but also reduces logistical challenges and travel times, thereby improving the overall impact of da'wah activities in Riau Province.

5. Conclusion

This study demonstrates the effectiveness of combining unsupervised learning algorithms with interactive geographic visualization to optimize the assignment of preachers in Riau Province. We have leveraged the unique advantages of K-means and DBSCAN algorithm to achieve more robust and precise clustering results.

The K-Means algorithm established a baseline for clustering, offering simplicity and higher Silhouette Scores. The optimal number of clusters was 10, with a Silhouette Score of 0.435654. However, K-Means inability to handle outliers was evident, as indicated by zero outliers detected across all configurations. In contrast, the DBSCAN algorithm effectively managed outliers and varying densities, providing a more flexible clustering solution. The optimal DBSCAN configuration with an epsilon of 1.5 and 5 minimum samples resulted in 10 clusters, a Silhouette Score of 0.381108 and identified 60 outliers. This study demonstrated DBSCAN's fine performance in handling noise and non-spherical clusters.

Integrating these clustering algorithms with interactive geographic visualization tools such as Folium has enabled the thorough exploration and analysis of spatial data. The interactive maps created in this study provided an intuitive and comprehensive view of the distribution of preachers and mosques, aiding in the decision-making process for preacher assignments. This approach addressed logistical challenges and ensured religious teachings were more relevant and impactful for different communities.

Our study can significantly contribute to this case study, mainly because it discusses the problem of outlier handling and comprehensively conducts a comparative analysis of clustering algorithms with appropriate parameters. This study enhances the robustness and accuracy of clustering results, thereby improving the efficiency and effectiveness of preacher deployment. The framework developed in this study is expected to be adapted and implemented in other regions to solve religious problems and contribute more widely.

Future studies could explore integrating other machine learning techniques and further refine the algorithms to handle even more complex datasets, ultimately enhancing the impact of preaching activities. This study will evolve beyond a model to produce a web and mobile-based system that will map preaching activities in Riau Province, aptly named the Smart *Da'wah* Map.

Acknowledgement

We sincerely thank the Institute of Research and Community Services (LPPM), Universitas Riau, for funding this research under Grant No. 15666/UN19.5.1.3/AL.04/2024, which enabled its successful completion. We also acknowledge the valuable support provided by the Indonesian Council of Ulama, Riau Province (MUI Riau).

References

Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics*, *9*(8), 1295. https://doi.org/10.3390/electronics9081295

Balavand, A. (2022). *Combination of a New Metaheuristic Algorithm Based on Cooperative Grouper Fish - Octopus and DBSCAN Algorithm to Automatic Clustering*. https://doi.org/10.21203/rs.3.rs-1118095/v1

Cintia Ganesha Putri, D., Leu, J.-S., & Seda, P. (2020). Design of an Unsupervised Machine Learning-Based Movie Recommender System. *Symmetry*, *12*(2), 185. https://doi.org/10.3390/sym12020185

Dibia, V., & Demiralp, C. (2019). Data2Vis: Automatic Generation of Data Visualizations Using Sequence-to-Sequence Recurrent Neural Networks. *IEEE Computer Graphics and Applications*, *39*(5), 33–46. https://doi.org/10.1109/MCG.2019.2924636

Fernandes, F., Correia, J., & Pontes, A. (2023). Business Intelligence: Trends and the Impact of Emerging Technologies. *Iberian Conference on Information Systems and Technologies, CISTI*, *2023-June*, 1–6. https://doi.org/10.23919/CISTI58278.2023.10211645

HabiBoğlu, M. G., Hernandez, H. W., & Uyaver, Ş. (2022). Shape investigations of structures formed by the self-assembly of aromatic amino acids using the density-based spatial clustering of applications with noise algorithm. *Turkish Journal of Electrical Engineering and Computer Sciences*, *30*(1), 200–215. https://doi.org/10.3906/elk-2003-144

Hengki, Rizan, O., Adiwinoto, B., Supardi, Saputro, S. H., & Perkasa, E. B. (2021). Business Intelligence to Support Visualization of Indonesian Capital Market Investment Gallery Performance. *3rd International Conference on Cybernetics and Intelligent Systems, ICORIS 2021*. https://doi.org/10.1109/ICORIS52787.2021.9649610

Hu, H., Liu, J., Zhang, X., & Fang, M. (2023). An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis. *Pattern Recognition*, *139*, 109404. https://doi.org/10.1016/j.patcog.2023.109404

Islam, M. Z., Estivill-Castro, V., Rahman, M. A., & Bossomaier, T. (2018). Combining K-

Means and a genetic algorithm through a novel arrangement of genetic operators for high quality clustering. *Expert Systems with Applications*, *91*, 402–417. https://doi.org/10.1016/j.eswa.2017.09.005

Kurniawan, R., Abdullah, S. N. H. S., Lestari, F., Nazri, M. Z. A., Akhmad, M., & Adnan, N. (2020). Clustering and Correlation Methods for Predicting Coronavirus COVID-19 Risk Analysis in Pandemic Countries. *International Conference On Cyber & IT Service Management*, 138.

Kurniawan, R., Abdullah, S. N. H. S., Lestari, F., Nazri, M. Z. A., Mujahidin, A., & Adnan, N. (2020). Clustering and Correlation Methods for Predicting Coronavirus COVID-19 Risk Analysis in Pandemic Countries. *2020 8th International Conference on Cyber and IT Service Management, CITSM 2020*, 1–5. https://doi.org/10.1109/CITSM50537.2020.9268920

Lai, W., Zhou, M., Hu, F., Bian, K., & Song, Q. (2019). A New DBSCAN Parameters Determination Method Based on Improved MVO. *IEEE Access*, *7*, 104085–104095. https://doi.org/10.1109/ACCESS.2019.2931334

Lavalle, A., Maté, A., Trujillo, J., & Rizzi, S. (2019). Visualization requirements for business intelligence analytics: A goal-based, iterative framework. *Proceedings of the IEEE International Conference on Requirements Engineering*, *2019-September*, 109–119. https://doi.org/10.1109/RE.2019.00022

Monalisa, S., & Kurnia, F. (2019). Analysis of DBSCAN and K-means algorithm for evaluating outlier on RFM model of customer behaviour. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *17*(1), 110. https://doi.org/10.12928/telkomnika.v17i1.9394

Nur, A., Kurniawan, R., Nazri, M. Z. A., Rajab, K., Papilo, P., & Mas'ari, A. (2021). Solution to Traveling Freelance Teacher Problem using the Simple K-Means Clustering. *Proceedings - 2021 4th International Conference on Computer and Informatics Engineering: IT-Based Digital Industrial Innovation for the Welfare of Society, IC2IE 2021*, 112–116. https://doi.org/10.1109/IC2IE53219.2021.9649086

Otaviya, S. A., & Rani, L. N. (2020). Productivity and Its Determinants in Islamic Banks: Evidence From Indonesia. *Journal of Islamic Monetary Economics and Finance*, *6*(1), 189–212. https://doi.org/10.21098/jimf.v6i1.1146

Province, S. of R. (2024). *Population by Regency/City (Person), 2022-2023*. https://riau.bps.go.id/id/statistics-table/2/MzIjMg==/jumlah-penduduk-menurut-kabupaten-kota.html

Raja, M., Hasan, P., Mahmudunnobe, M., Saifuddin, M., & Hasan, S. N. (2024). Membership determination in open clusters using the DBSCAN Clustering Algorithm. *Astronomy and Computing*, *47*, 100826. https://doi.org/10.1016/j.ascom.2024.100826

Ran, X., Zhou, X., Lei, M., Tepsan, W., & Deng, W. (2021). A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots. *Applied Sciences*, *11*(23), 11202. https://doi.org/10.3390/app112311202

Sani, S. D., & Abubakar, M. (2021). A proposed framework for implementing risk-based Shari'ah audit. *Journal of Financial Reporting and Accounting*, *19*(3), 349–368. https://doi.org/10.1108/JFRA-02-2020-0041

Santosa, R. G., Lukito, Y., & Chrismanto, A. R. (2021). Classification and Prediction of Students' GPA Using K-Means Clustering Algorithm to Assist Student Admission Process. *Journal of Information Systems Engineering and Business Intelligence*, *7*(1), 1. https://doi.org/10.20473/jisebi.7.1.1-10

Tang, H., Deng, L., & Huang, Y. (2022). Business Intelligence System Based on Big Data Technology. *Proceedings - 2022 International Conference on Artificial Intelligence of Things and Crowdsensing, AIoTCs 2022*, 143–147. https://doi.org/10.1109/AIoTCs58181.2022.00027

Wahyuni, S. N., Khanom, N. N., & Astuti, Y. (2023). K-Means Algorithm Analysis for Election Cluster Prediction. *JOIV : International Journal on Informatics Visualization*, *7*(1), 1. https://doi.org/10.30630/joiv.7.1.1107

Wang, Q., Ming, Y., Jin, Z., Shen, Q., Liu, D., Smith, M. J., Veeramachaneni, K., & Qu, H.

(2019). ATMSeer. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300911

Wang, T., Ren, C., Luo, Y., & Tian, J. (2019). NS-DBSCAN: A Density-Based Clustering Algorithm in Network Space. *ISPRS International Journal of Geo-Information*, *8*(5), 218. https://doi.org/10.3390/ijgi8050218

Yang, Y., Qian, C., Li, H., Gao, Y., Wu, J., Liu, C., & Zhao, S. (2022). An efficient DBSCAN optimized by arithmetic optimization algorithm with opposition-based learning. *The Journal of Supercomputing*, *78*(18), 19566–19604. https://doi.org/10.1007/s11227-022-04634-w