Discovering Right Congregation For Preachers Using Clustering Approach

Afrizal Nur[1], Rahmad Kurniawan[2], Mohd Zakree Ahmad Nazri[3], Abdul Somad Batubara[4], Elin Haerani[5], Imansyah Putra[6]

[1] Faculty of Usuluddin, Universitas Islam Negeri Sultan Syarif Kasim Riau, 28293 Pekanbaru, Riau, Indonesia
[2,5] Department of Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Jl. HR. Subrantas Km. 15, Pekanbaru 28293, Indonesia
[3] Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
[4] Visiting Professor at Universiti Islam Sultan Sharif Ali, Brunei Darussalam
[6] RK Data Science Research Institute

Corresponding e-mail: rahmadkurniawan@uin-suska.ac.id

Abstract. Indonesia, as the most populous Muslim-majority country in the world, with more than 500,000 mosques and prayer rooms, each with its congregation. Religious lectures are delivered by preachers who have been selected and appointed by the Mosque management. A mosque is functioning as the center for Islamic Education has its unique method in determining preacher, but preachers too need a way to know the mosque's congregation location. Mosques are located both in urban areas, where many mosques are nearby and at remote villages, which are sparsely populated rural areas, where the travel time between mosques is significant. The preacher’s problem can be seen as the travelling sales man with time window problem. Therefore, this study aims to develop a mosque recommender system for preachers to find the nearest mosque using machine learning. Machine learning uses informatics and statistical was deemed capable of optimizing the preacher's schedule. In this study, we used the K-means to automatically search for unknown clusters of mosques that are similar to a preacher's geolocation. Results show that by using K-Means, the approach can recommend similar congregation (mosque) to the preacher's requirement at an accuracy of about 83%. This study also revealed the best number of clusters were determined by cluster number index, cluster sum of squared errors, silhouette and time taken.

Keywords: Clustering, K-means, machine learning, Muslim, schedule

1. Introduction

Indonesia is the most populous Muslim-majority country in the world. For Muslims, the mosque is the central place for worship. Currently, there are 270,253 Mosques and 316,468 prayer rooms throughout Indonesia. In Riau Province alone, there are 6,610 mosques and 877 mosques in Pekanbaru City [1]. Besides as a prayer places, a mosque is also functioning as the center for Islamic education and *da'wah*. The word "da'wah" (Arabic دعوة) is basic to a study of Islam. Mosques have been the epicenter for religious and cultural in a Muslim community. It has social, political, and judicial functions. The mosque congregations come from the nearby communities residents, and therefore their religious understandings and principles can be strongly influenced by the community. A spiritual preacher mission for da'wah is always looking for ways to maximize the impact of their lectures on society. However, the location and time of the given lecture are another two important factors for selecting a congregation.

Sometimes, a preacher will be offered as an invited speaker or teacher from multiple mosques or apply to a various mosque at the same time to give a lecture. It is usual for preachers to flood many mosques with numerous applications. The problem is they are not taking the time to study and individualize their applications to show that they are both qualified to fill the particular slot but also can arrive on time to the mosque an any given time and date.

Every Mosque in Indonesia has a regular schedule for Islamic lectures with a strict time window. A preacher does not only have to select the right congregation, but also the correct location so that a preacher can arrive at the mosque within the given schedules. Mosques are located both in urban and village areas. Many mosques in the urban areas are in close proximity. Still, mosques are far from each other in the sparsely populated rural areas, where the travel time between mosques is significant. Mosque management is also facing the same problems in determining the right preacher from thousands of applicants and candidates for guest speakers. Often, a selected or invited preacher cancelled their given lecture slot because of the location of the congregation is far from their early lectures or residence or temporary accommodation. Therefore, it is necessary to increase and improve the *Da'wah* management system for mosques.

In this paper, the preacher's dilemma can be seen as the famous computer science problem, which is the *travelling salesman problem with time windows*. Thus, the preacher's problem can be seen as a problem that involves in finding the minimum cost tour in which all mosques are visited once or more than once within their requested time windows. This optimization problem has numerous important practical applications, including scheduling and routing. This problem can be regarded as NP-complete, and hence traditional optimization algorithms have not been applied to solve the preacher's problem. However, the discovery of the right congregation places involves obtaining the physical locations and their similarity to the preacher's profile that matters to both, believe and expertise.

The rapid enlargement of technology and computers have been employed in various fields including in Islamic *Da'wah* [2–6]. Nevertheless, the literature review showed that no research assists Mosque managements in determining preacher based on their location distance. Hence, clustering of mosques using the accurate algorithms to determine preacher based on location is necessary. This study proposed a machine learning approach for discovering the most appropriate congregation location for a preacher. As a case study for this research, we observe mosques which are managed by an Islamic organization called Muhammadiyah. Muhammadiyah is one of a popular Islamic organization in Indonesia which has built and managing mosques throughout Indonesia. There are 32 Muhammadiyah-manage mosques and prayer rooms in Pekanbaru City.

Machine learning that uses informatics and statistical analysis can solve the problem, which is also commonly used to solve specific problems that are qualified. We used a generally unsupervised machine learning technique, namely clustering, in this study. This well-known method of clustering is capable of agglomerating unlabeled information regardless of the number of attributes. It is also suitable to be used on this underlying open data [7], namely Mosques data consist of longitude and latitude [1].

The K-means clustering algorithm is selected extensively to determine the preacher in the Muhammadiyah Mosques because of its efficacy and simplicity. A recent study used the K-means algorithm to clustering 25 mammals [8]. Although it used small data, it has produced excellent clustering results [8,9]. The study also compared various methods to determine the best number of clustering. It indicated that the lowest amount of clusters was right [8,9]. Therefore, this study employs K-means clustering for clustering Muhammadiyah Mosques in Pekanbaru. This study also aims to observe the essential parameters to validate the performance and consistency of the cluster results.

2. Material and Methods

Data were obtained from three sources, namely from Muhammadiyah Pekanbaru, SIMAS [1] and Google maps [10]. The data consists of three attributes, namely the Mosque names, latitudes and longitudes. We obtained 877 coordinates of Mosques in Pekanbaru. However, to make the research more focused, 32 Muhammadiyah Mosques and 86 Preachers were clustered.

2.1. Data Pre-processing

We obtained 800 Mosques data in Pekanbaru. Based on the data were provided by Muhammadiyah, 32 Mosque selected as Muhammadiyah Mosques. Furthermore, 400 Preachers in Riau Province data were obtained then 86 of them are domiciled in Pekanbaru were chosen to be processed.

2.2. Clustering Model

Eventually, the K-means algorithm to cluster the Muhammadiyah Mosques in Pekanbaru. Let the Euclidean distance between β_s and γ_t, δ_s is the number Muhammadiyah Mosques in Pekanbaru coordinate points at the cluster, and δ is the number of cluster centers, K-means clustering algorithm satisfies $\alpha(\beta, \gamma, \delta)$, an objective function minimization of squared error function $\|\beta_s - \gamma_t\|$. Given that $\beta = \{\beta_1, \ldots, \beta_n\}; n = 3$ be the set of Muhammadiyah Mosques in Pekanbaru coordinate points comprising of *Mosque names, latitudes and longitudes* and $\delta = \{\delta_1, .., \delta_m\}$ be the set of center points, and the number of the cluster that tested are $m = \{5,10,15,20,30\}$ subsequently. After defining n, step 1 randomly select δ cluster centers. Then, step 2 estimates the distance between each Mosques data point, β_s and cluster centers, γ_t. Upon completion, step 3 substitutes the Mosques data point, β_s to the cluster center, γ_t that constitutes the lowest distance, α of the cluster center compared to all the cluster centers, δ. Next, it re-estimates the new cluster center with the average value of $\gamma_s = \frac{1}{\delta_s}\sum_{t=1}^{\delta_s} \beta_s$, and distance between each Mosques data point, β_s and the new cluster centers, γ_{new} at Step 4 and 5 subsequently. At final step 6, it will stop the whole process if no Mosques data point was substituted else proceed to Step 3.

2.3. Testing and Evaluation

Several parameters were selected as experiment settings, i.e. *distance function, initialization method, max iteration, the total number of items, number of clusters, cluster number index (performance), within-cluster sum of squared errors* and *time is taken.*

3. Results and Discussion

In this section, we run five experiments, as shown in Table I. We have used the simple K-means algorithm for clustering Muhammadiyah Mosques in Pekanbaru.

Table 1. Experiment settings of the K-means clustering algorithm

Experiment settings	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**
Distance function	Euclidean distance	Euclidean distance	Euclidean distance	Euclidean distance	Euclidean distance
Initialization method	Random	Random	Random	Random	Random
Max iteration	300	300	300	300	300
Total number of items (Mosques and Preachers)	118	118	118	118	118
Number of clusters, *m*	5	10	15	20	30
Cluster number index (Performance)	83%	65%	43%	31%	3%
Within cluster sum of squared errors *(α)*	24	19	14	9	1
Average of silhouette	0.65	0.61	0.6	0.59	0.51
Time taken	0.01 seconds	0.01 seconds	0.01 seconds	0.01 seconds	0.01 seconds

Based on Table 1, it achieved the best cluster number index (performance) when the number of clusters m=5. Therefore, the experiment of clustering Muhammadiyah Mosques was conducted when the number of clusters m=5. Table 2 shows the generated silhouette score to calculate how close a Mosque is to a cluster of its own compared with other clusters. The silhouette close to 1 for a given data instance means that the data instance is close to the middle of the cluster. Instances with silhouette scores close to 0 are located between two clusters on the border.

Table 2. The example results of clustering ten Mosques when the number of clusters m=5, α=24 at Run 1

No.	Mosques / Prayer Rooms	Clusters	Silhouette Scores
1.	*M. Taqwa Ps. Pusat*	C1	0.538963
2.	*M. Taqwa Srikandi*	C1	0.54119
3.	*M. Al-Furqan*	C4	0.512022
4.	*M.Taqwa Cikpuan*	C1	0.687439
5.	*M.Taqwa Cipta Karya*	C3	0.671883
6.	*Ms.Al-Ikhlas (Panam)*	C3	0.712008
7.	*Ms.Al Khusaini*	C3	0.707142
8.	*Ms.Nurul Jamik*	C3	0.713341
9.	*Ms.Jabal Nur*	C2	0.684333
10.	*Ms.Nurul Iman*	C5	0.576911

Based on Table 2, it is inferred that the Mosques clustering are acceptable as the data instance is close to the centre of the cluster. Subsequently, an experiment was to organize the data of Preacher and Mosques. Suppose Mosque management aimed to know which preacher would be assigned by location to a Mosque. We also confirmed the results obtained by the K-Means clustering algorithm by looking at the actual distance on the Google maps. The map in Figure 3*(b)* shows that a Preacher was successfully assigned at the nearest mosque, which is 2.6 km from his resident.

4. Conclusion

This research tries to find remedies for the preacher's problem. We applied a clustering algorithm to discover congregations. We also used an essential evaluation metric as part of the evaluation framework. We then conducted an experiment that collected real location data and illustrated the utility of our evaluation framework. This study accomplished using the K-Means algorithm to cluster Muhammadiyah Mosques in Pekanbaru. The Mosque clustering aims to facilitate the decision that a preacher has to make so that they can arrive at the mosque based on location. Also, this study revealed that the number of clusters was determined by cluster number index (performance), cluster sum of squared errors, silhouette scores and time is taken. The K-Means algorithm performs correctly and fasts computational. Furthermore, based on the silhouette scores, it was found that the small number of clusters lead to consistent clusters and best performance. This study is still in the early experimental stage. Our results establish a baseline that future work can measure itself against. They also demonstrate that our algorithm discovers places with reasonable accuracy. Our next step in this research is to integrate the clusters with the routing optimization algorithm.

Acknowledgments

Universitas Islam Negeri Sultan Syarif Kasim Riau, Universiti Kebangsaan Malaysia, RK Data Science Research Institute and Muhammadiyah in Riau support this work.

References

[1] Indonesian Ministry of Religious Affairs, 2020, “Sistem Informasi Masjid” [Online]. Available: https://simas.kemenag.go.id/. [Accessed: 03-Nov-2020].

[2] Kurniawan, R., Akbarizan, Jamal, K., Nur, A., Ahmad, M. Z., and Kholilah, D., 2018, “Advise-Giving Expert Systems Based on Islamic Jurisprudence for Treating Drugs and Substance Abuse,” J. Theor. Appl. Inf. Technol.

[3] Jamal, K., Kurniawan, R., Batubara, A. S., Nazri, M. Z. A., Lestari, F., and Papilo, P., 2020, “Text Classification on Islamic Jurisprudence Using Machine Learning Techniques,” *Journal of Physics: Conference Series*, Institute of Physics Publishing, p. 12066.

[4] Ashaari, M. F., and Ghani, R. A., 2010, “The Model of Online Al-Dacwah Al-Fardiyyah,” *Proceeding of the 3rd International Conference on Information and Communication Technology for the Moslem World: ICT Connecting Cultures, ICT4M 2010*.

[5] Draman, M. Z., and Zeki, A. M., 2014, “3D Animation as an Alternative Mean of Da’wah for Children,” *2014 the 5th International Conference on Information and Communication Technology for the Muslim World, ICT4M 2014*, Institute of Electrical and Electronics Engineers Inc.

[6] Murtada, A., and Mansor, A., 2016, “Mosque Finding and Mobile Profile Changing Application,” *Proceedings - 2015 International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering, ICCNEEE 2015*, Institute of Electrical and Electronics Engineers Inc., pp. 485–490.

[7] Baruri, R., Ghosh, A., Banerjee, R., Das, A., Mandal, A., and Halder, T., 2019, “An Empirical Evaluation of K-Means Clustering Technique and Comparison,” Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Prespectives Prospect. Com. 2019, pp. 470–475.

[8] Matt, 2019, “10 Tips for Choosing the Optimal Number of Clusters,” Towar. Data Sci. [Online]. Available: https://towardsdatascience.com/10-tips-for-choosing-the-optimal-number-of-clusters-277e93d72d92. [Accessed: 16-Apr-2020].

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

[10] Google Inc., 2020, “Pekanbaru - Google Maps” [Online]. Available: https://www.google.com/maps/place/Pekanbaru,+Pekanbaru+City,+Riau/@0.5137908,101.3711346,12z/data=!3m1!4b1!4m5!3m4!1s0x31d5ab80690ee7b1:0x94dde92c3823dbe4!8m2!3d0.5070677!4d101.4477793. [Accessed: 05-Nov-2020].

[11] Bhardwaj, A., 2020, “Silhouette Coefficient Validating Clustering Techniques,” Towar. Data Sience [Online]. Available: https://towardsdatascience.com/silhouette-coefficient-validating-clustering-techniques-e976bb81d10c. [Accessed: 26-Nov-2020].