

Data Clusterization of Muslim Majority Countries to Find Out the Most Factors Causing Gender Issues Using the K-Means Algorithm

Rifqi Syamsul Fuadi
Department of Informatics
UIN Sunan Gunung Djati
Bandung, Indonesia
rifqi@uinsgd.ac.id

Dian Sa'adillah Maylawati
Department of Informatics
UIN Sunan Gunung Djati
Bandung, Indonesia
diansm@uinsgd.ac.id

Ramdhan Nugraha Pratama
Department of Informatics
UIN Sunan Gunung Djati
Bandung, Indonesia
ramdhannp6@gmail.com

Akhdan Musyaffa Firdaus
Department of Informatics
UIN Sunan Gunung Djati
Bandung, Indonesia
1197050010@student.uinsgd.ac.id

Dea Puminda Sari
Department of Informatics
UIN Sunan Gunung Djati
Bandung, Indonesia
1197050027@student.uinsgd.ac.id

Maulana Hamdani
Department of Informatics
UIN Sunan Gunung Djati
Bandung, Indonesia
1197050066@student.uinsgd.ac.id

Novia Nurhanivah
Department of Informatics
UIN Sunan Gunung Djati
Bandung, Indonesia
1197050098@student.uinsgd.ac.id

Abstract— Islam is a religion that upholds the dignity of women. Even Islam teaches that heaven is at the feet of a mother. However, in countries where most of the population is Muslim, gender issues persist. Therefore, this article was created to know the factors that most contribute to gender issues in various countries where most of the population is Muslim. These factors are divided into internal factors and external factors. Internal factors include those that are relevant to oneself. External factors include things other than oneself, such as culture. The author uses the K-Means algorithm as the algorithm used to manage the retrieved data. The author uses a collection of survey data found on the Internet about the reactions of men and women to Islam and gender issues in Muslim-dominated countries.

Keywords—gender issues, K-Means algorithm, Muslim-dominated countries

I. INTRODUCTION

In Islamic teachings, men and women have the same position and opportunity to act in social life [1]. In principle, Islam does not prohibit women from working inside or outside the home independently or in collaboration with the private sector or the government, day or night, as long as the work is carried out with respect and can carry out religious obligations. Advice and can avoid the negative impact of his work on himself, his family, and the environment [2].

Islam also does not impose certain hours and certain precautions for work. He emphasized that work should not be too heavy a burden because of the hours worked and the nature of the work [3]. However, based on UNDP research, the gender inequality index remains high in countries with Muslim majority populations, such as Indonesia, Bangladesh, Pakistan, Turkey, Uzbekistan, and others. Looking at the inequality or gender issues that occur in this Muslim-majority country, we collect them by looking for the factors that cause gender problems and two characteristics that cause problems, namely gender problems caused by internal factors and external factors [4].

Internal factors exist within women themselves. For example, women do not dare to express their opinions or prefer traditional life roles. External factors, namely factors outside of women themselves, and the most dominant is the existence of patriarchal cultural values that regulate all life in family society. Therefore the role of women is secondary [5], [6].

As a result, the inequality or gender issues that occur in Muslim-majority countries become interesting to study and examine.

Based on the background of the problem described, it will be obtained the main issues that will be discussed in this study. The question is how to apply the K-Means algorithm to classify the factors that cause gender problems in Muslim-majority countries.

II. METHODS

This research phase begins with a preliminary study. This is the first step in research by collecting and researching relevant data to gain insight into the ongoing investigation. In addition, the method of collecting and processing data in this study is data collection which is done by searching for datasets on the internet. Data collection techniques are also used in the form of library research. We collect relevant data, information, and theory from the literature to support research analysis.

Clustering is a phenomenon used in data science to group similar data objects. It is primarily a search function to identify similar and different items and group them accordingly [7]. Data clustering is widely used in data mining and pattern recognition. There are several methods to extend the cluster, such as the hierarchical method, partition method, grid-based method, model-based method, etc. The k-means clustering algorithm is a widely used method for clustering data. The main goal of this algorithm is to minimize the distance between each data point and the center of its cluster. Figure 1 further illustrates the basic diagram of the k-means algorithm's elbow method and presents its schematic.

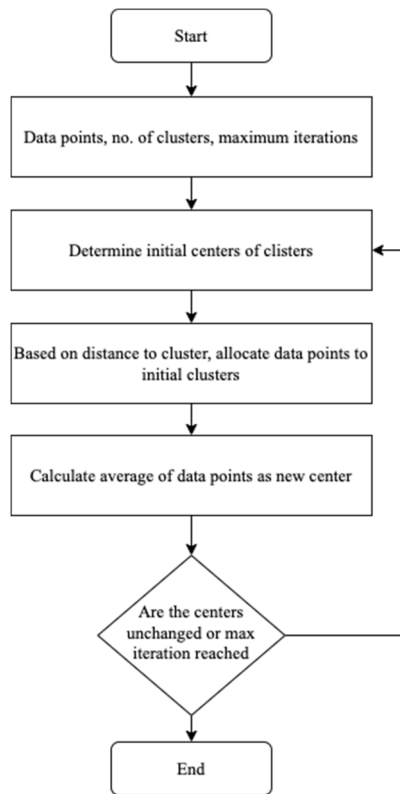


Fig. 1. Flowchart of the k-means clustering algorithm.

The grouping of data in this study was the K-Means method. K-Means clustering is a method of collecting or classifying a set of objects with the same attributes or characteristics into several groups [5], [6], [8], [9]. The reason the author uses K-Means is that this method can be used if we group data based on several variables where we cannot determine what the output class is (unsupervised learning) [10]–[12]. In addition, this algorithm can be used for problems whose solution requires a segmentation process or grouping into specific subgroups. We can also use the k-means clustering algorithm during Exploratory Data Analysis (EDA) to complete descriptive statistical analysis and data visualization [13].

In this study, we are interested in finding the factors that cause gender problems. The results of this study do not describe the number of gender problems that occur but only the general factors that cause gender problems in Muslim countries. Survey data have different characteristics. Survey data we divide into two groups, then predict in the document to determine the group of internal factors and groups of external factors [5], [6], [14].

Survey questions:

- survey 1: currently plays a significant role (Islam in politics)
- survey 2: Should play a considerable part (Islam in politics)
- survey 3: Democracy can work here
- survey 4: Women should be permitted to work outside the home
- survey 5: Husband provides for the family, and the wife cares for the house and children
- survey 6: Husband and wife both work

III. RESULT AND DISCUSSION

This study aims to find out the most common factors that cause gender problems in various Muslim countries. We will divide these factors into two, namely internal factors and external factors. Internal factors include things related to oneself. At the same time, extrinsic factors include things outside the self, such as culture. We conducted this study using the K-Means algorithm. This study was conducted using a dataset of surveys found on the Internet about the reactions of men and women to gender and Islam in a country with a sizeable Muslim population.

In general, this research includes the following discussions:

- Select the variables that will be the reference for conducting the analysis and give reasons.
- Identify the methodology used to analyze and explain why this method is chosen (state and briefly describe the theory).
- I am displaying the data exploration results and interpreting the analysis results.
- At the end of the discussion, recommendations are made based on the research results.

A. Dataset

The data set used in this study includes the results of a survey on male and female responses to Islam and gender issues in Muslim-majority countries [15]. There are 13 variables, namely:

- country: name of Muslim-majority countries
- survey1_men: male survey results in % (1)
- survey1_women: survey results of women in % (1)
- survey2_men: male survey results in % (2)
- survey2_women: women survey results in % (2)
- survey3_men: male survey results in % (3)
- survey3_women: survey results of women in % (3)
- survey4_men: male survey results in % (4)
- survey4_women: survey results of women in % (4)
- survey5_men: male survey results in % (5)
- survey5_women: survey results of women in % (5)
- survey6_men: male survey results in % (6)
- survey6_women: survey results of women in % (6)

The use of the dataset can be seen in Table 1 below:

TABLE I. SURVEY RESULT DATASET

Country	survey1_men	survey1_women	survey2_men	survey2_women	survey3_men	survey3_women	survey4_men	survey4_women	survey5_men	survey5_women	survey6_men	survey6_women
Bangladesh	53	51	83	85	76	42	36	57	52	50	47	51
Indonesia	87	86	84	82	43	38	20	24	49	42	51	58
Jordan	58	48	85	66	66	71	13	16	58	67	41	33
Lebanon	65	65	64	62	72	69	58	75	42	32	56	68
Mali	63	60	73	66	79	73	51	56	44	30	55	70
Nigeria	70	85	66	79	76	75	32	38	31	28	69	72
Pakistan	52	64	91	85	63	51	24	41	58	57	31	38
Senegal	65	64	46	42	87	87	60	68	23	11	77	78
Tanzania	16	13	15	22	67	61	47	47	34	26	63	74
Turkey	45	45	40	44	51	59	60	72	36	23	63	75

B. Determining the Number of Clusters

The number of clusters corresponds to the number of buckets to be created. One way is to use the elbow method based on the number of variations calculated from several K values.

From the visual results in Figure 2, it will be challenging to determine the number of K that fits the data. Therefore, bend analysis will be carried out to determine the optimal number of K for the data used.

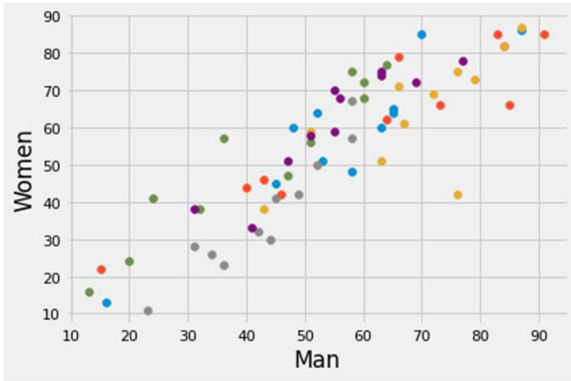


Fig. 2. The result of seeing the number of variations in calculation

Based on Figure 3, it can be seen that at point 2, there is a rather steep bend. Therefore we can conclude that $K = 2$ is the optimal value.

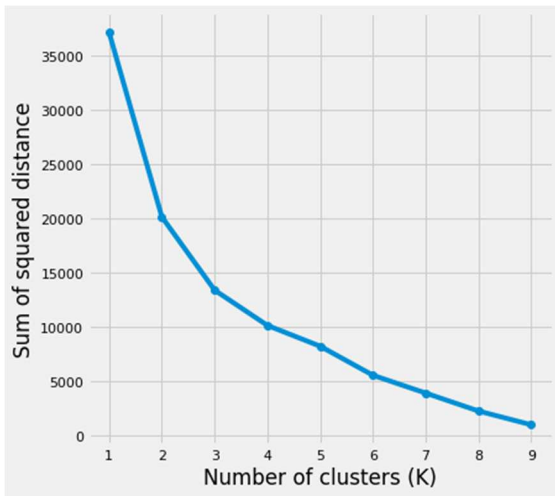


Fig. 3. Sum squared distance

C. Implementation of K-Means Clustering

First, create a group for each survey, then identify or label the group as internal and external factors. This is done because each survey has a different conclusion. Not every study that gets many vote is an external factor, and vice versa. After conducting a group survey, it is necessary to determine which group is internal and external.

Figure 4 is the result of survey cluster 1 obtained from the data from Table 1, more precisely Survey1_men and Survey1_women.

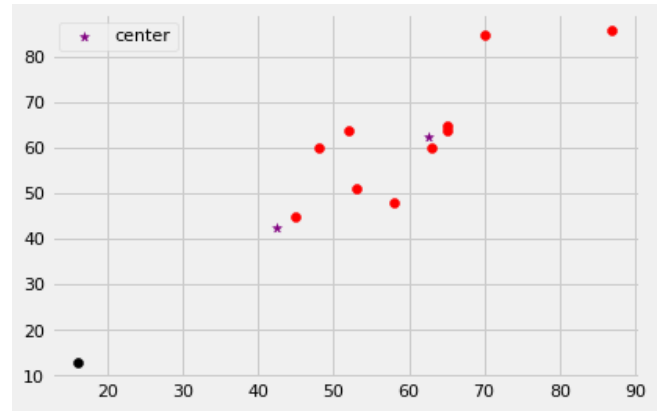


Fig. 4. The results of survey cluster 1

Figure 5 is the result of survey cluster 2 obtained from the data from Table 1, more precisely Survey2_men and Survey2_women.

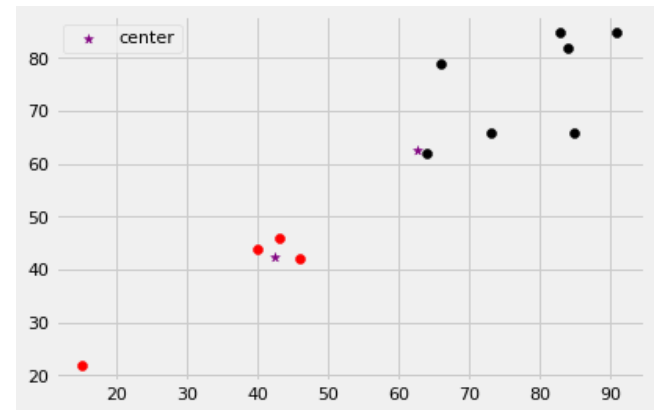


Fig. 5. The result of survey cluster 2

Figure 6 is the result of survey cluster 3 obtained from the data from Table 1, more precisely Survey3_men and Survey3_women.

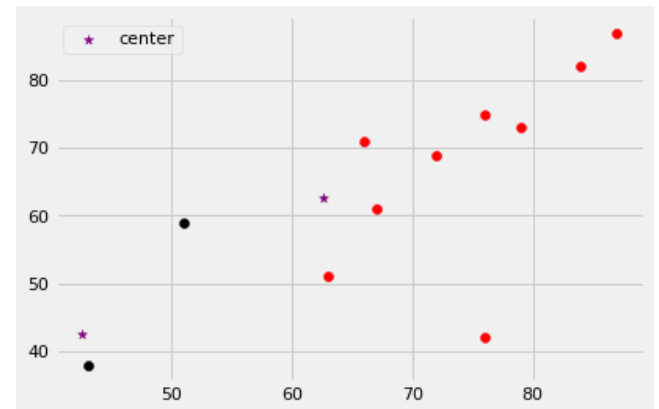


Fig. 6. The result of survey cluster 3

Figure 7 is the result of survey cluster 4 obtained from the data from Table 1, more precisely Survey4_men and Survey4_women.

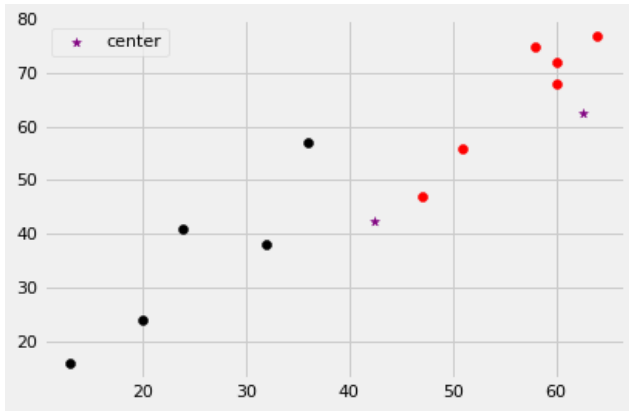


Fig. 7. The result of survey cluster 4

Figure 8 is the result of survey cluster 5 obtained from the data from Table 1, more precisely Survey5_{men} and Survey5_{women}.

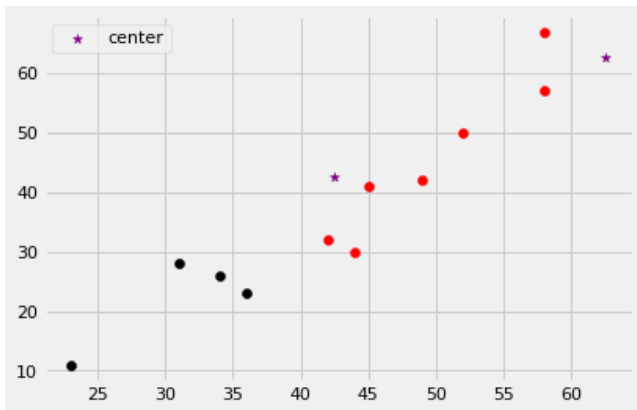


Fig. 8. The result of survey cluster 5

Figure 9 shows survey cluster 6 obtained from Table 1 data, more precisely Survey6_{men} and Survey6_{women}.

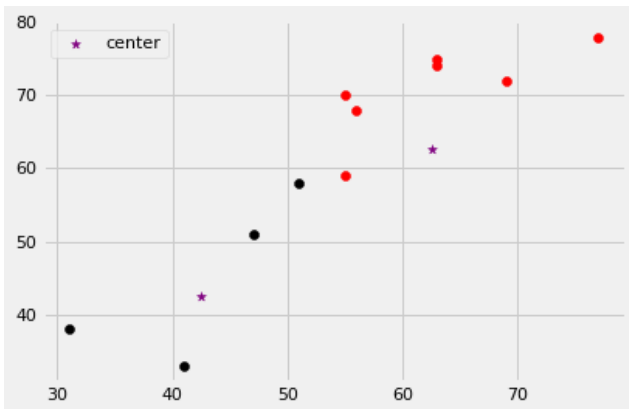


Fig. 9. The result of survey cluster 6

Cluster results in each survey can be seen in Table II.

TABLE II. RESULTS IN EACH SURVEY

country	survey_1	survey_2	survey_3	survey_4	survey_5	survey_6
Bangladesh	0	0	0	0	1	0
Indonesia	0	0	1	0	1	0
Jordan	0	0	0	0	1	0
Lebanon	0	0	0	1	1	0
Mali	0	0	0	1	1	0
Nigeria	0	0	0	0	0	1
Pakistan	0	0	0	0	1	0
Senegal	0	1	0	1	0	1
Tanzania	1	1	0	1	0	1
Turkey	0	1	1	1	0	1
Uzbekistan	0	1	0	1	1	0

After clustering each survey, then define and label each cluster. The results for the groups that have been labeled can be seen in Table III.

TABLE III. LABELED CLUSTER RESULTS

country	survey_1	survey_2	survey_3	survey_4	survey_5	survey_6
Bangladesh	eksternal	internal	internal	eksternal	eksternal	eksternal
Indonesia	eksternal	internal	eksternal	eksternal	eksternal	eksternal
Jordan	eksternal	internal	internal	eksternal	eksternal	eksternal
Lebanon	eksternal	internal	internal	internal	eksternal	eksternal
Mali	eksternal	internal	internal	internal	eksternal	eksternal
Nigeria	eksternal	internal	internal	eksternal	internal	internal
Pakistan	eksternal	internal	internal	eksternal	eksternal	eksternal
Senegal	eksternal	eksternal	internal	internal	internal	internal
Tanzania	internal	eksternal	internal	internal	internal	internal
Turkey	eksternal	eksternal	eksternal	internal	internal	internal
Uzbekistan	eksternal	eksternal	internal	internal	eksternal	eksternal

The final results are obtained from the survey data clustered to determine the factors that most cause gender issues in various countries where most people are Muslim. These results are presented in Figure 10 below.

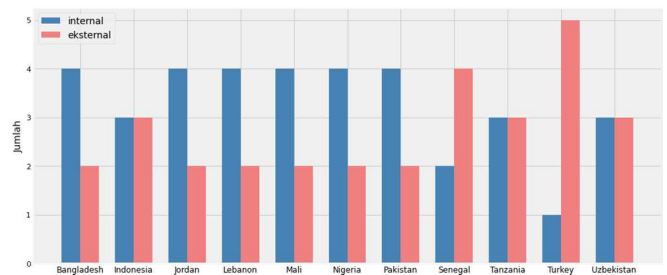


Fig. 10. the country with the most gender issues

Figure 10 shows that the countries where gender issues arise the most due to extrinsic factors are Turkey and Senegal, while for Bangladesh, Jordan, Lebanon, Mali, and Nigeria, the most common factor originating from the root of gender problems is internal factors. And for Indonesia, Tanzania, and Uzbekistan, there are similar results.

IV. CONCLUSION

The K-Means method can be applied to synthesize many factors that contribute to gender issues in Muslim-majority countries. The result of this application is that the most common factors causing gender problems in most countries are internal factors, such as in Bangladesh, Jordan, Lebanon, Mali, Nigeria, and Pakistan. However, there are also countries whose gender problems are caused by external factors, such as Senegal and Turkey. Meanwhile, Indonesia and Tanzania achieved a balance between internal and external factors.

In the final results, there are similarities in results in several countries because the attributes used are only six

surveys, so the probability of producing a consistent final result is very high.

ACKNOWLEDGMENT

The authors wish to acknowledge the Research and Publication Centre of UIN Sunan Gunung Djati Bandung, which supports and funds this research publication.

REFERENCES

- [1] "Gender Equity in Islam." [Online]. Available: <https://www.iium.edu.my/deed/articles/genderequityinislam.html>. [Accessed: 14-Mar-2022].
- [2] A. Kia, "The concept of responsibility of men and women in Islam," *Arts Humanit. Open Access J.*, vol. Volume 3, no. Issue 5, Oct. 2019.
- [3] G. Catolino, F. Palomba, D. A. Tamburri, A. Serebrenik, and F. Ferrucci, "Gender Diversity and Community Smells: Insights from the Trenches," *IEEE Softw.*, vol. 37, no. 1, pp. 10–16, Jan. 2020.
- [4] W. Ben-Amar, M. Chang, and P. McIlkenny, "Board Gender Diversity and Corporate Response to Sustainability Initiatives: Evidence from the Carbon Disclosure Project," *J. Bus. Ethics*, vol. 142, no. 2, pp. 369–383, May 2017.
- [5] J. Wu, J. Chen, H. Xiong, and M. Xie, "External validation measures for K-means clustering: A data distribution perspective," *Expert Syst. Appl.*, vol. 36, no. 3 PART 2, pp. 6050–6061, 2009.
- [6] L. Jegatha Deborah, R. Baskaran, and A. Kannan, "A Survey on Internal Validity Measure for Cluster Validation," *Int. J. Comput. Sci. Eng. Surv.*, vol. 1, no. 2, pp. 85–102, Nov. 2010.
- [7] S. Kapil and M. Chawla, "Performance evaluation of K-means clustering algorithm with various distance metrics," undefined, Feb. 2016.
- [8] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *J. Intell. Inf. Syst.*, vol. 17, no. 2–3, pp. 107–145, Dec. 2001.
- [9] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010.
- [10] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020.
- [11] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J*, vol. 2, no. 2, pp. 226–235, Jun. 2019.
- [12] I. V. Stuldreher, A. Merasli, N. Thammasan, J. B. F. van Erp, and A.-M. Brouwer, "Unsupervised Clustering of Individuals Sharing Selective Attentional Focus Using Physiological Synchrony," *Front. Neuroergonomics*, vol. 2, Jan. 2022.
- [13] A. Mexicano, J. C. Carmona, S. Cervantes, J. A. Cervantes, S. López, and R. Rodríguez, "A Modified Version of K-Means Algorithm," *Lect. Notes Networks Syst.*, vol. 343, pp. 299–308, 2022.
- [14] N. D. and B. M. Vidyavathi, "A Survey on Applications of Data Mining using Clustering Techniques," *Int. J. Comput. Appl.*, vol. 126, no. 2, pp. 7–12, Sep. 2015.
- [15] "Global Gender Gaps | Pew Research Center." [Online]. Available: <https://www.pewresearch.org/global/2004/05/13/global-gender-gaps-2/>. [Accessed: 30-Apr-2022].