JOIV : Int. J. Inform. Visualization, 7(3) - September 2023 977-983

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

K-Means Clustering Algorithm for Partitioning the Openness Levels of Open Government Data Portals

Emigawaty [a,b,*], Kusworo Adi [a], Adian Fatchur Rochim [c], Budi Warsito [a], Adi Wibowo [a]

[a] *Doctoral Program of Information System, Universitas Diponegoro, Tembalang, Semarang, 50275, Indonesia*
[b] *Department of Informatics, Universitas Amikom, Sleman, Yogyakarta, 55281, Indonesia*
[c] *Department of Computer Engineering, Universitas Diponegoro, Tembalang, Semarang, 50275, Indonesia*
*Corresponding author: *emigawaty@amikom.ac.id*

***Abstract*— More and more local governments in Indonesia are making their data available to the public. This benefits data scientists, researchers, business owners, and other potential users seeking datasets for empirical research and business innovation. However, just because Open Government Data (OGD) portals are accessible does not mean that they necessarily adhere to the established rules and principles of data openness. To evaluate the level of openness of 24 OGD portals in Indonesia, this study used the K-means Clustering algorithm to partition them into three levels: Leaders, Followers, and Beginners. A group of 30 participants, including researchers, data scientists, business enablers, and graduate students, rated the portals on 32 sub-questions related to the eight main principles of data disclosure, focusing on health, population, and education datasets. The study found that eight portals were categorized as Leaders, ten as Followers, and seven as Beginners regarding their level of openness. The study demonstrated that the K-means Clustering algorithm can be effectively used to assess the degree of openness of OGD portals in Indonesia based on eight main principles of data openness. The study recommends increasing the number of OGD portals in eastern territories to supplement the existing case studies in the western and central regions.**

***Keywords*— K-Means; clustering; open government data; portals.**

Manuscript received 20 Apr. 2023; revised 19 Jun. 2023; accepted 31 Jul. 2023. Date of publication 10 Sep. 2023.

I. INTRODUCTION

Opening up government datasets can positively impact citizen engagement and increase trust in the government by promoting transparency and accountability [1]–[5]. An open data portal is a website platform that provides public access to data in the public sector that can be reused [6]–[8]. Open Government Data (OGD) portals are web-based interfaces that make it easier for people to access information that can be repurposed [9]–[12]. There are numerous benefits to disclosing data through OGD portals, including generating new knowledge, promoting transparency and accountability, receiving up-to-date information on government ideas and achievements, fostering business innovation, supporting policies and decisions, and other possible value propositions [13]–[15].

The OGD initiative was introduced in Indonesia in 2010 in response to the need for greater transparency and accountability in the country's governance. The initiative was primarily aimed at making government data and information more accessible to the public, promoting public participation, fostering innovation and economic growth, and enhancing overall government performance.

This initiative was made possible by establishing a legal framework provided by Law (UU) Number 14 of 2008 about Public Information Disclosure (KIP), which mandated the government to disclose any information related to the public interest, including data, documents, and other information. In order to effectively implement the UU KIP in Indonesia, the government issued Government Regulation (PP) Number 61 of 2010, which provided specific guidelines for data disclosure, including standards for data management, processing, and dissemination. This regulation also mandated the creation of Open Government Data portals to facilitate public access to government data and information and promote greater transparency and accountability in governance.

However, conducting regular evaluations of OGD portals at the national and regional levels in Indonesia is important to ensure that the data they provide meets established standards.

Despite the rise in the number of data portals launched by local governments in Indonesia, which offers a new and valuable source of information for the public, data users, including researchers and commercial innovators, there is no guarantee that these portals are complying with established regulations or the eight fundamental principles that serve as guidelines for data disclosure.

The present study, which involved a systematic literature review, revealed a lack of evidence regarding earlier research on OGD portals. Previous qualitative and quantitative studies that have examined the success rate of OGD portals may have yielded inconsistent results, possibly due to different assessment models and an inability to meet the diverse needs of various stakeholders who use data [15], [16]. The present study employs a mixed-method approach to address the concern about discrepancies in previous research on OGD portals.

Given these circumstances, this study aims to evaluate 24 OGD portals in Indonesia and classify their degree of openness into three distinct categories: leader, follower, and beginner [17], [18]. The assessment was based on the eight fundamental principles of data disclosure established by a group of 30 open government advocates in Sebastopol, California, at the end of 2007 [19], [20]. These principles include completeness, primacy, timeliness, accessibility, machine-readability, non-discrimination, non-proprietary, and licensing-free [6], [10], [21], [22].

Our approach to achieving this objective involves utilizing the K-means clustering algorithm to cluster OGD portals in Indonesia. Furthermore, we assembled a diverse and representative sample of 30 participants comprising researchers, data scientists, business enablers, and university graduate students. These participants were asked to respond to 32 sub-questions related to the eight primary principles of data disclosure based on their firsthand experience with OGD portals and to rate each question on a scale of one to five. Regarding data classification, we grouped the datasets into three categories: health, population, and education.

II. Material and Method

A. Dataset and Experimental Survey

This section discusses how datasets were selected from each OGD portal. For each portal, three categories of datasets were chosen to ensure that the structure and value of the data had similar characteristics. This was important to ensure that the evaluation of the experimental study could be balanced across the datasets held in each region, whether at the provincial or district and city level in Indonesia. The datasets were classified into three categories: health, population, and education. The experimental survey instrument consisted of eight primary questions and 32 sub-questions, presented in detail in Table 1.

This study utilizes various techniques to assess the level of openness of OGD portals in Indonesia. The methods employed can be broadly categorized into two main parts. Firstly, an experimental survey instrument was developed based on the eight open data principles. Participants were asked to rate the questions on a scale of one to five, drawing from their experience with OGD portals. This approach ensured that the evaluation of the questions was grounded in practical experiences rather than just theoretical knowledge.

TABLE I
DESIGN OF EXPERIMENTAL SURVEY INSTRUMENT

	Principle	Variable	Question
1	Complete	P1	Is all public data available? (data that has no valid privacy, security, or privilege limits is considered public data)
		P1.1	Is the dataset complete in terms of properties and fields?
		P1.2	Is the dataset complete in terms of data content?
		P1.3	Is it possible to download all the records from the dataset?
		P1.4	Is the dataset re-usable after it has been downloaded?
2	Primary	P2	Is the data collected from the source or data owner (department, ministry, government agency) at the maximum level of detail available without being aggregated or modified?
		P2.1	Is the dataset from a well-known data source?
		P2.2	Are the datasets available original and unaltered?
		P2.3	Do the themes and categories of the datasets match the available ones?
		P2.4	Is the available dataset the data provider's first source?
3	Timely	P3	Is the data readily available and up-to-date, as is required to sustain the data's value?
		P3.1	Is it possible to access and download the dataset quickly?
		P3.2	Is the available dataset currently renewable (for a maximum of one year)?
		P3.3	Is there any information about the reference date in the dataset?
		P3.4	Is there any information in the dataset about when it was developed and/or made available on the portal?
4	Accessible	P4	Is the information accessible to a wide range of users for a wide range of purposes?
		P4.1	Is the available dataset easy to find using the portal's existing web navigation system?
		P4.2	Is the available dataset usable on a variety of devices?
		P4.3	Are the offered datasets accessible to people of various ages and genders?
		P4.4	Do people with particular requirements (blindness, etc.) easily access the available datasets?
5	Machine-processible	P5	Is the data sufficiently structured (metadata) to allow for automated processing?

	Principle	Variable	Question
		P5.1	Is it possible for computer systems and other devices to read and reuse the dataset?
		P5.2	Is the dataset available in various file formats, such as CSV, XLS, XML, JSON, and so on?
		P5.3	Is there a text format for the dataset that is easy to read?
		P5.4	Is it possible to use the dataset for data analysis and programming?
6	Non-discriminatory	P6	Is the information accessible to anyone without the need to register?
		P6.1	Is the dataset freely available to anyone who wants to use it?
		P6.2	Is it true that accessing the dataset through the portal does not require authentication?
		P6.3	Does the portal distinguish in terms of user classification when it comes to gaining access to the dataset?
		P6.4	Is the anonymization of dataset user access on the portal system guaranteed?
7	Non-proprietary	P7	Is the data in a format that no one entity has exclusive control over?
		P7.1	Are there no dataset ownership labels on the datasets that are available?
		P7.2	Is it true that specific people have exclusive access to the available dataset portals?
		P7.3	Is it possible to access and reuse the existing datasets in the future?
		P7.4	Are the datasets available in various forms and sources of data ownership accessible and downloadable?
8	License-free	P8	Are there any copyright, patent, trademark, or trade secret regulations the data is aware of? (License-free, subject to acceptable privacy, security, and privilege constraints.)
		P8.1	Is the data you have access to copyrighted?
		P8.2	Is the dataset available trademarked by any company or organization?
		P8.3	Is there an explanation for privacy, copyright, and licenses in the available datasets?
		P8.4	Is there anything in the available dataset that indicates it can only be accessed by non-public?

To evaluate the selected OGD portals, we used an experimental survey questionnaire and invited 30 users of OGD portals to participate in the evaluation process. The participants were drawn from various stakeholders in open data and divided into four groups: 10 researchers and lecturers, 7 data scientists, 4 business enablers, and 9 university graduate students who volunteered to participate. The evaluation was conducted on 24 OGD portals between November 2022 and February 2023.

B. K-Mean Algorithm

In addition, to understand how the K-means Clustering method works, we followed these steps [23], [24]: (1) specify the number of clusters (K) that the algorithm will create; (2) randomly select K data points and assign each to a cluster based on the number of data points in each category; (3) calculate the cluster centroids; (4) repeat the process until the data points are assigned to stable clusters.

We used the Euclidean distance method to measure the distance in the clustering observations. This method calculates the distance between each data point and the initial cluster centroids 1, 2, and 3. The Euclidean distance method aims to measure the distance between two data points in a multidimensional space. It calculates the straight-line distance between two points, considering each dimension as a coordinate in space [25].

The formulation of the Euclidean can be illustrated as follows:

$$\sqrt{(X_H - H_1)}^{\,2} + (X_W - H_1)^{\,2} \qquad (1)$$
$$\sqrt{(X_H - H_1)}^{\,2} + (X_W - H_1)^{\,2}$$

where,

$X_H X_H$: Observation value of variable Height

$H_1 H_1$: Centroid value of Cluster 1 for available Height

$X_W X_W$: Observation value of variable Weight

$W_1 W_1$: Centroid value of Cluster 1 for variable Weight

After generating the K-means Clustering, we use the Google Colabs platform and the matplotlib library to create a 3D scatterplot representing each OGD level's clusters.

III. Results and Discussion

Following the experimental surveys conducted in this study, several significant findings have emerged concerning applying the K-means Clustering method to categorize the level of openness of OGD portals. These findings have been classified into three main categories, which are presented in detail in Table 2. Firstly, the study ascertained the overall level of OGD portals in local Indonesian governments based on the experience and knowledge of the participants. Secondly, the study employed the K-means Clustering method to determine the degree of openness of the portals. Lastly, the Google Colabs platform was leveraged to create a 3D scatterplot that effectively summarizes the level of openness of each OGD portal. These findings highlight the efficacy of the K-means Clustering method for assessing the openness of OGD portals and provide insights into the current state of data openness in local Indonesian governments.

A. Experimental Survey

Table 2 reflects the mapping results of assessing the quality of OGD portals in 24 provinces, districts, and cities in Indonesia. The city of Mataram obtained the highest total

score from the survey conducted for the eight-question criteria designed in Table 1, followed by three other cities: Mojokerto, Denpasar, and Pare-Pare. Meanwhile, the illustration the ranking of the achievement of the eight OGD openness criteria can be seen in Figure 1.

TABLE II
RESULTS OF THE EXPERIMENTAL SURVEY IN THIS STUDY

	OGD Portal	URL	Health Dataset	Population Dataset	Education Dataset	Score Total	Score Average
1	Mataram	https://data.mataramkota.go.id/	47.70	48.97	47.41	144.07	48.02
2	Mojokerto	http://data.mojokertokota.go.id/	44.09	47.25	46.25	137.59	45.86
3	Denpasar	https://pusatdata.denpasarkota.go.id/	45.09	45.41	45.41	135.91	45.30
4	Pare-pare	https://satudata.pareparekota.go.id/	45.22	45.22	45.22	135.66	45.22
5	Medan	https://data.pemkomedan.go.id/	44.66	44.97	44.91	134.53	44.84
6	Cirebon	https://data.cirebonkota.go.id/	46.13	45.44	42.75	134.31	44.77
7	Kota Batu	https://portaldata.batukota.go.id/	44.94	44.94	43.94	133.81	44.60
8	Jakarta	https://data.jakarta.go.id/	44.55	44.56	44.03	133.15	44.38
9	Banjarmasin	https://data.kalselprov.go.id/	44.25	44.25	44.09	132.59	44.20
10	Pekalongan	https://data.pekalongankota.go.id/	43.97	43.97	43.97	131.91	43.97
11	Bekasi	https://danta.bekasikota.go.id/	42.88	43.84	43.28	130.00	43.33
12	Pontianak	https://data.pontianakkota.go.id/	43.03	43.03	43.66	129.72	43.24
13	Tasikmalaya	https://data.tasikmalayakota.go.id/	40.72	43.94	44.88	129.53	43.18
14	Bima	https://data.bimakota.go.id/	42.81	43.75	42.78	129.34	43.11
15	Bandung	http://data.bandung.go.id/	42.19	43.84	43.09	129.13	43.04
16	Magelang	https://data.magelangkota.go.id/	42.59	43.13	43.13	128.84	42.95
17	Dumai	https://opendata.dumaikota.go.id/	40.66	40.66	45.53	126.84	42.28
18	Cimahi	https://data.cimahikota.go.id/	42.13	42.35	42.35	126.83	42.28
19	Tegal	http://data.tegalkab.go.id/	41.72	41.72	41.41	124.84	41.61
20	Bogor	https://data.kotabogor.go.id/	42.47	39.84	37.97	120.28	40.09
21	Jambi	http://data.jambikota.go.id/	32.19	35.31	44.16	111.66	37.22
22	Malang	https://opendata.malangkota.go.id/	37.06	34.56	36.03	107.66	35.89
23	Metro	https://data.metrokota.go.id/	33.41	33.41	33.41	100.22	33.41
24	Sukabumi	https://opendata.sukabumikota.go.id/	33.22	33.22	33.22	99.66	33.22

Following that, we can visualize the usefulness and openness of the OGD portal based on the participants' perspectives and experiences in analyzing the three types of datasets, as shown in Figure 1.

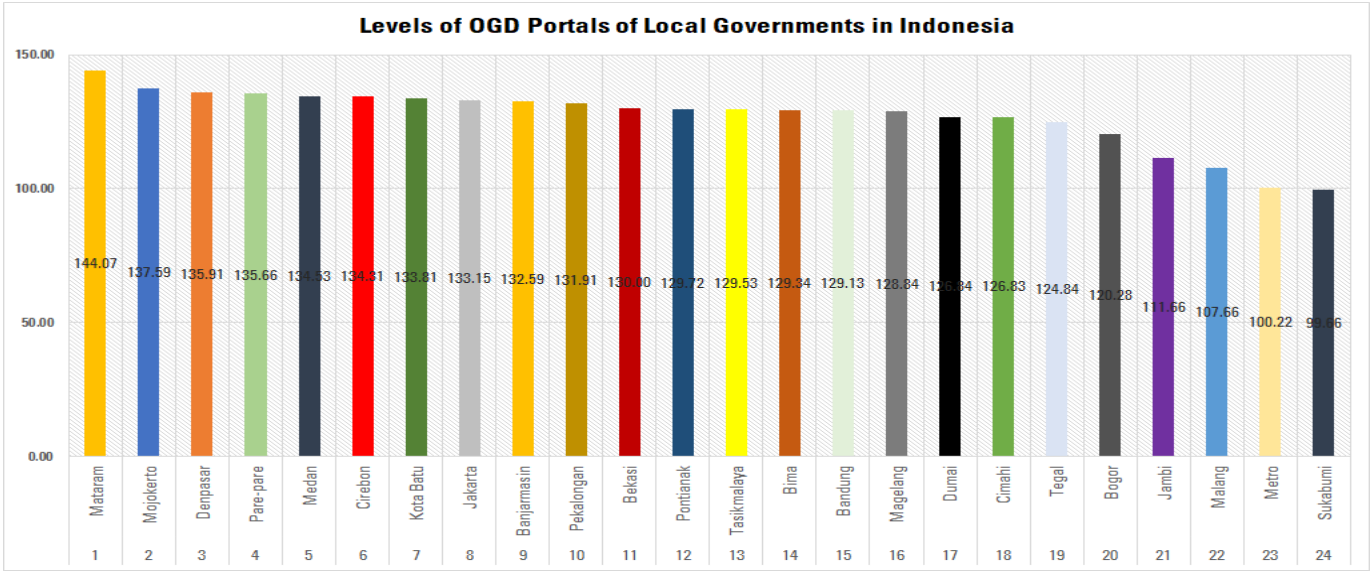

Fig. 1 Levels of OGD Portals Resulted from the Experiment Surveys

B. K-Means Clustering Algorithm

The study visualized the usefulness and openness of the OGD portal using Figure 1, which is based on the participants' experiences and perspectives of analyzing three different types of datasets. Then, the K-Means Clustering Algorithm technique was applied to determine the nearest distance and optimal cluster based on the experimental findings of the OGD portal assessment by selected participants. The determination of centroid value is considered a significant requirement for implementing the K-Means method.

Utilizing the K-means Clustering method, the study identified three initial centroid locations for the three distinct clusters of OGD portals. Specifically, Jakarta City was chosen

as the centroid for Cluster 1, Banjarmasin City was designated as the second centroid, Tegal City was selected as the third centroid, and conditioned for Cluster 3. The distance between each data point and the previously selected centroid was utilized to determine the new centroid, with a new centroid being chosen from the data points with a frequency equivalent to its distance from the prior centroid. The outcomes of the OGD portal partitioning at the level of the K-means Clustering method are presented in detail in Table 3.

TABLE III
RESULTS OF THE K-MEAN CLUSTERING ALGORITHM

	OGD Portal	Health Dataset	Population Dataset	Education Dataset	Cluster 1	Cluster 2	Cluster 3	Nearest Cluster	Ideal Cluster
1	Mataram	47.70	48.97	47.41	33.95	36.69	94.54	33.95	Cluster 1
2	Mojokerto	44.09	47.25	46.25	12.60	13.81	56.43	12.60	Cluster 1
3	Denpasar	45.09	45.41	45.41	3.14	3.90	32.97	3.14	Cluster 1
4	Pare-pare	45.22	45.22	45.22	2.51	3.17	30.29	2.51	Cluster 1
5	Medan	44.66	44.97	44.91	1.03	1.58	25.75	1.03	Cluster 1
6	Cirebon	46.13	45.44	42.75	3.98	5.09	20.04	3.98	Cluster 1
7	Kota Batu	44.94	44.94	43.94	0.53	1.18	19.99	0.53	Cluster 1
8	Jakarta	44.55	44.56	44.03	0.00	0.40	17.81	0.00	Cluster 1
9	Banjarmasin	44.25	44.25	44.09	0.40	0.00	16.16	0.00	Cluster 2
10	Pekalongan	43.97	43.97	43.97	0.94	0.38	13.88	0.38	Cluster 2
11	Bekasi	42.88	43.84	43.28	2.76	2.20	9.19	2.20	Cluster 2
12	Pontianak	43.03	43.03	43.66	4.01	2.90	8.10	2.90	Cluster 2
13	Tasikmalaya	40.72	43.94	44.88	4.94	4.24	17.96	4.24	Cluster 2
14	Bima	42.81	43.75	42.78	3.96	3.41	7.11	3.41	Cluster 2
15	Bandung	42.19	43.84	43.09	3.76	3.23	7.83	3.23	Cluster 2
16	Magelang	42.59	43.13	43.13	4.85	3.86	5.81	3.86	Cluster 2
17	Dumai	40.66	40.66	45.53	21.40	18.58	19.21	18.58	Cluster 2
18	Cimahi	42.13	42.35	42.35	10.12	8.75	1.71	1.71	Cluster 3
19	Tegal	41.72	41.72	41.41	17.81	16.16	0.00	0.00	Cluster 3
20	Bogor	42.47	39.84	37.97	61.10	58.71	16.08	16.08	Cluster 3
21	Jambi	32.19	35.31	44.16	97.94	91.95	58.13	58.13	Cluster 3
22	Malang	37.06	34.56	36.03	171.49	166.04	84.76	84.76	Cluster 3
23	Metro	33.41	33.41	33.41	248.50	242.65	141.41	141.41	Cluster 3
24	Sukabumi	33.22	33.22	33.22	256.92	250.99	147.79	147.79	Cluster 3

C. *Levels of OGD Portals*

In the next step, we utilized Google Colabs to depict the level of success in terms of openness of 24 OGD portals belonging to various local governments and cities in Indonesia. This was achieved by creating a scatter plot graph, which visualizes the results of applying the K-Means Clustering algorithm to these portals, dividing them into three clusters. The scatter plot presents data points on two axes, showing the relationship between two variables [23], [26], [27]. The X and Y axes indicate the location of each data point [24], [28], [29]. We also provide some code snippets that can be used to create a scatter plot on Google Colabs.

```
# Import library
import matplotlib.pyplot as plt
import pandas as pd
import pylab as pl
import numpy as np
%matplotlib inline
import seaborn as sns
df = pd.read_csv('/Dataset_OGD.csv') #lokasi file
csv
df.head(3)
# Import OGD Portal Survey dataset
dataset = pd.read_csv('/Dataset_OGD.csv')
X = dataset.iloc[:,2:5].values
X[:3]
# Implement K-Means Clustering into dataset with 3
clusters
kmeans = KMeans(n_clusters = 3, init = 'k-means++',
random_state = 42)
y_kmeans = kmeans.fit_predict(X)
dataset['f_cluster'] = y_kmeans + 1
dataset.head(3)
# 3d scatterplot using matplotlib
fig = plt.figure(figsize = (6,6))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(X[y_kmeans==0,0],X[y_kmeans==0,1],X[y_km
eans==0,2],s=100,color='deeppink',label = "Cluster
1")
ax.scatter(X[y_kmeans==1,0],X[y_kmeans==1,1],X[y_km
eans==1,2],s=100,color='orange', label = "Cluster
2")
ax.scatter(X[y_kmeans==2,0],X[y_kmeans==2,1],X[y_km
eans==2,2],s=100,color='dodgerblue', label =
"Cluster 3")
ax.set_xlabel('Beginners')
ax.set_ylabel('Followers')
ax.set_zlabel('Leaders')
ax.legend()
plt.show()
```

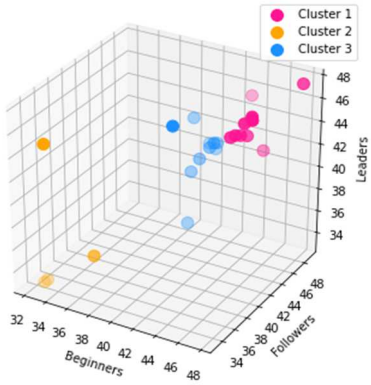

Fig. 2 Levels of OGD Portals Resulted from the Experiment Surveys

Figure 2 shows the data openness proportion of 24 OGD portals in Indonesia using a Scatter Plot graph. The visualization model provided by Google Colabs helps to identify which OGD portals are in Cluster 1 (Leader level), Cluster 2 (Follower level), and Cluster 3 (Beginning level). Figure 3 represents the partition of the openness level of the 24 OGD clusters based on the three developed clusters. Cluster 1 (Leader level) includes 8 OGD portals: Mataram, Mojokerto, Denpasar, Pare-pare, Medan, Cirebon, Kota Batu, and Jakarta. Cluster 2 (Follower level) includes 9 OGD portals: Banjarmasin, Pekalongan, Bekasi, Pontianak, Tasikmalaya, Bima, Bandung, Magelang, and Dumai. Cluster 3 (Beginner status) includes 7 OGD portals: Cimahi, Tegal, Bogor, Jambi, Malang, Metro, and Sukabumi. Based on their openness, the K-Means Clustering technique has successfully split the OGD portals into three different levels.

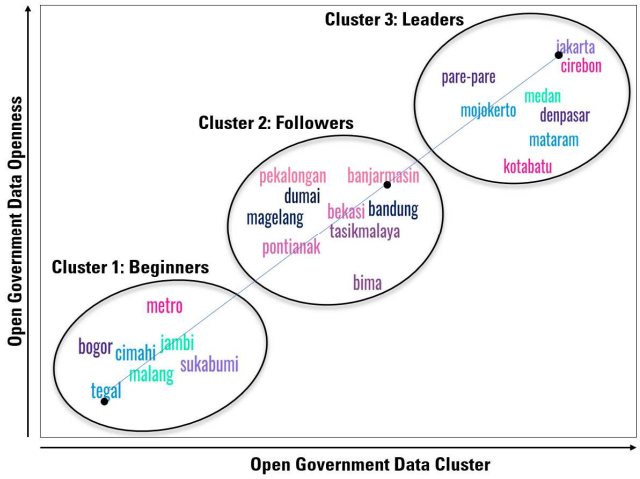

Fig. 3 Levels of OGD Portals in terms of Openness

Researchers are inquisitive about the success of the OGD portal's implementation in Indonesia as a result of this study. The success indicator is how far the local government has designed and set the OGD portals in compliance with globally formulated data disclosure regulations or standards. As a result, this research employs eight key data-disclosure principles developed by a group of 30 open-government advocates in Sebastopol, California, at the end of 2007. Several index models, like ODRA, G-ODI, and ODB, employed these standards to evaluate the success of OGD portals.

Given the difficulties in collecting potential volunteers through the OGD portal, the study only included 30 people to evaluate the success of obtaining data from 24 OGD portals across Indonesia. Because they come from a variety of backgrounds, including researchers and lecturers, data scientists, business people, and students, these participant representatives can be used as legitimate samples in general. However, we encourage increasing the number of respondents from OGD portal users with a variety of backgrounds, such as journalists, non-governmental organizations, active consumers, ordinary residents, and even within the local government itself, in the next research. As a result, the findings of experimental analysis related to the OGD portal's adoption in Indonesia become more reliable and exhaustive.

Conversely, we comprehensively assess the K-means Clustering algorithm's potential shortcomings. The most significant drawback of K-means Clustering is that the value of K (number of clusters) must be determined at the beginning of the process [30]. This means that k-means can only deal with numbers. On the other hand, K-means assumes that we have been dealing with clusters with roughly the same number of observations in each [31], [32]. In this study, the value of K is determined using the common classification of data openness levels in general, namely Leaders, Followers, and Beginners.

A few key findings in the study might be explored while designing additional research. First, the results of an experimental survey with multiple participants still reflect the varied and subjective opinions of the participants based on their degree of education and experience. This subjectivity will undoubtedly impact the position of the open data portal cluster. Therefore, several academic and governmental professionals must validate the survey results to lessen the likelihood of this bias in the following research. With the adoption of OGD in Indonesia, there is a gap between survey respondents' perceptions and the actual situation. This check and balance seek to close that gap. Second, the distribution of representation of the Indonesian territory based on geographical and regional administrative places is considered in this study's selection of OGD portal samples. As a result, there is a chance that the OGD sample selection will not be accepted and may not be balanced between the OGD's readiness at the provincial level and smaller administrative areas. As a result, by segmenting the assessment area into sub-regions, the following study can employ a clustered distribution model that is more precise.

IV. Conclusion

In this study, the K-means clustering algorithm was utilized to categorize the levels of OGD portal openness in Indonesia based on eight key data openness principles. The research findings indicate that eight OGD portals belong to Cluster 1, which is classified as the leader level. These portals have advanced technological capabilities, innovative features, and sophisticated functionalities that facilitate cross-domain coordination. Meanwhile, Cluster 2 is comprised of 9 OGD portals classified as the follower level, which provide basic features such as direct visibility and progressive performance characteristics. However, their data release approach is fragmented, and data availability is limited. The remaining OGD portals have been assigned to Cluster 3 with a Beginner level, indicating that local governments are taking the first steps but struggling to meet availability and accessibility criteria. Moreover, the OGD portal's functionality is limited at this level, and the dataset coverage is insufficient.

In order to expand upon the findings presented in this study, future research can explore additional dimensions of Open Government Data (OGD) portals, such as the quality and relevance of the data provided, the user experience, and the impact of OGD on public participation and decision-making processes. Additionally, future studies can extend the analysis to other regions or countries to compare the level of openness of OGD portals and identify best practices for promoting data transparency and accountability. Furthermore, incorporating other clustering algorithms or applying

different distance metrics can be considered to examine the clustering results' robustness and identify alternative OGD portal categorizations.

REFERENCES

[1] A. Luthfi dan M. Janssen, "A conceptual model of decision-making support for opening data," in *Communications in Computer and Information Science*, 2017, vol. 792, doi: 10.1007/978-3-319-71117-1_7.

[2] S. de Juana-Espinosa dan S. Luján-Mora, *de Juana-Espinosa, S., & Luján-Mora, S. (2020). Open government data portals in the European Union: A dataset from 2015 to 2017*, vol. 29. 2020.

[3] M. Malacaria, S., De Mauro, A., Greco, M., & Grimaldi, "An Application of the Analytic Hierarchy Process to the Evaluation of Companies' Data Maturity," *ICEIS*, vol. 1, hal. 50–61, 2022.

[4] X. Zhu dan M. A. Freeman, "An evaluation of US municipal open data portals: A user interaction framework," *J. Assoc. Inf. Sci. Technol.*, vol. 70, no. 1, 2019, doi: 10.1002/asi.24081.

[5] M. Wen., "Leading successful government-academia collaborations using FLOSS and agile values," *J. Syst. Softw.*, vol. 164, 2020, doi: 10.1016/j.jss.2020.110548.

[6] Y. Charalabidis, A. Zuiderwijk, C. Alexopoulos, M. Janssen, T. Lampoltshammer, dan E. Ferro, *The World of Open Data: Concepts, Methods, Tools and Experiences*. Granada: Springer International, 2018.

[7] O. Belkahla Driss, S. Mellouli, dan Z. Trabelsi, "From citizens to government policy-makers: Social media data analysis," *Gov. Inf. Q.*, vol. 36, no. 3, 2019, doi: 10.1016/j.giq.2019.05.002.

[8] Y. Lixin, X., & Li, "Research on Maturity Assessment of Local Open Government Data Portals," *Libr. Inf. Serv.*, vol. 63, no. 12, 2019.

[9] L. Ayre, "Open Data: What It Is and Why You Should Care," *Jim Craner*, vol. 36, no. 2, hal. 173–184, 2017, doi: 10.1080/01616846.2017.1313045.

[10] D. Matheus, R., Janssen, M., & Maheshwari, "Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities," *Gov. Inf. Q.*, vol. 37, no. 3, 2022.

[11] B. B. Ansari, "Enhancing the usability and usefulness of open government data: A comprehensive review of the state of open government data visualization research," *Gov. Inf. Q.*, vol. 39, no. 4, 2021, doi: 10.1016/j.giq.2021.101657.

[12] A. Luthfi dan M. Janssen, "Toward a Reference Architecture for User-Oriented Open Government Data Portals," in *International Symposium on Business Modeling and Software Design*, 2022, hal. 259–267, doi: 10.1007/978-3-031-11510-3_17.

[13] T. Jetzek, M. Avital, dan N. Bjorn-Andersen, "The Sustainable Value of Open Government Data," *J. Assoc. Inf. Syst.*, vol. 20, no. 6, hal. 702–734, 2019, doi:10.17705/1jais.00549.

[14] A. Abella, M. Ortiz-de-Urbina-Criado, dan C. De-Pablos-Heredero, "Criteria for the identification of ineffective open data portals: pretender open data portals," *Prof. la Inf.*, vol. 31, no. 1, 2022, doi: 10.3145/epi.2022.ene.11.

[15] R. Máchová, M. Hub, dan M. Lněnička, "Usability evaluation of open data portals: Evaluating data discoverability, accessibility, and reusability from a stakeholders' perspective," *Aslib J. Inf. Manag.*, vol. 70, no. 4, 2018, doi: 10.1108/AJIM-02-2018-0026.

[16] G. M.Begany, E. G.Martin, dan X. (Jenny) Yuan, "Open government data portals: Predictors of site engagement among early users of Health Data NY," *Gov. Inf. Q.*, vol. 38, no. 4, 2021, doi: 10.1016/j.giq.2021.101614.

[17] T. Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, "Data governance: Organizing data for trustworthy Artificial Intelligence," *Gov. Inf. Q.*, vol. 37, no. 3, 2020.

[18] A. Twizeyimana, J. D., & Andersson, "The public value of E-Government–A literature review," *Gov. Inf. Q.*, vol. 36, no. 2, hal. 167–178, 2019.

[19] R. Pereira, G. V., Parycek, P., Falco, E., & Kleinhans, "Smart governance in the context of smart cities:A literature review," *Inf. Polity*, vol. 23, no. 2, hal. 143–162, 2018.

[20] P. Glyptis, L., Christofi, M., Vrontis, D., Del Giudice, M., Dimitriou, S., & Michael, "E-Government implementation challenges in small countries: The project manager's perspective," *Technol. Forecast. Soc. Change*, vol. 152, 2020.

[21] A. Quarati, M. De Martino, dan S. Rosim, "Geospatial open data usage and metadata quality," *J. Inf. Sci.*, 2021, [Daring]. Tersedia pada: https://www.mdpi.com/959152.

[22] N. Janssen, M., & Helbig, "Innovating and changing the policy-cycle: Policy-makers be prepared," *Gov. Inf. Q.*, vol. 35, no. 4, hal. 99–105, 2018.

[23] V. Govender, P., & Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980-2019)," *Atmos. Pollut. Res.*, vol. 11, no. 1, hal. 40–56, 2020.

[24] S. Fränti, P., & Sieranoja, "K-means properties on six clustering benchmark datasets," *Appl. Intell.*, vol. 48, hal. 4743–4759, 2018.

[25] Anne-Laure Mention, *Digital Innovation: Harnessing the Value of Open Data (Open Innovation: Bridging Theory and Practice)*. WSPC, 2019.

[26] S. M. S. Ahmed, M., Seraj, R., & Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, 2020.

[27] F. A. Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. D. F., & Rodrigues, "Clustering algorithms: A comparative approach," *PLoS One*, vol. 14, no. 1, 2019.

[28] S. S. Irfiani, E., & Rani, "Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita," *J. Sist. dan Teknol. Inf,* vol. 6, no. 4, hal. 161, 2018.

[29] A. Bouyer, A., & Hatamlou, "An efficient hybrid clustering method based on improved cuckoo optimization and modified particle swarm optimization algorithms," *Appl. Soft Comput.*, vol. 67, hal. 172–182, 2018.

[30] A. Fahim, "K and starting means for k-means algorithm," *J. Comput. Sci.*, vol. 55, 2021, doi: Journal of Computational Science.

[31] M. J. Rezaee, M. Eshkevari, M. Saberi, dan O. Hussain, "GBK-means clustering algorithm: An improvement to the K-means algorithm based on the bargaining game," *Knowledge-Based Syst.*, vol. 213, 2021, doi: 10.1016/j.knosys.2020.106672.

[32] Y. A. Elnaga dan S. Nasr, "K-means cluster interactive algorithm-based evolutionary approach for solving bilevel multi-objective programming problems," *Alexandria Eng. J.*, vol. 61, no. 1, hal. 811–827, 2022, doi: 10.1016/j.aej.2021.04.098.