Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 175–182
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50267

Clustering Analysis of E-Learning Readiness in Java Island Indonesia with GIS Visualization

Eva Khudzaeva[1*], Qurrotul Aini[2], Evy Nurmiati[3], Ismi Ana Sulasiyah[4], Ibrahim Shehu Usman[5]

***Abstract*—E-learning readiness (e-readiness) is used as a tool to measure the success rate of using ICT in the academic process. Until now, a lot of research on e-readiness has been collected in various regions, especially Java Island, but the data has not been grouped and visualized, so it is difficult to know the level of readiness. The purpose of this study is to group e-readiness data using clustering analysis, then create a GIS-based map of the distribution of e-readiness clusters. To obtain optimal clustering results, researchers used the K-means and PCA combination as a cluster optimization method. The total dataset used is 27 locations' data with 2 parameters selected based on the level of e-learning readiness. Based on the results of the performance analysis using the selected internal clustering validation metrics, specifically the Davies Bouldin Index (DBI) and the average within centroid distance, each metric indicates the best cluster with values of 0.057 and 0.001, respectively. The most optimal cluster formed using the K-means and PCA methods, with a total of three clusters spread across various areas on the island of Java. As for the division of each cluster by its location point, namely, cluster 1 (amounting to 20 locations) for the ready level, cluster 2 (amounting to 4 locations) for the less ready level, and cluster 3 (amounting to 3 locations) for the very ready level.**

***Index Terms*—E-learning readiness, clustering analysis, K-means, principal component analysis, visualization, GIS.**

I. INTRODUCTION

A system that uses e-learning is different from previous learning systems, learning using e-learning requires preparation both in terms of infrastructure and the technical abilities of the users involved. To measure the level of success in using information technology in the learning process, a tool is needed to assess e-learning readiness in the learning process [1]. In Indonesia, quite a lot of research data on e-learning readiness (e-readiness) has been collected which is published freely on the internet, but the data has not been grouped and visualized so it is difficult to know the level of readiness in each region. E-readiness research data published on the internet contains information on regional areas, academic levels, samples that become research variables, values and categories of user readiness, and the year the research was conducted. According to the literature study obtained by the researcher, Java Island was chosen because 81% of research on e-readiness from 2013–2021 was conducted there. The success of e-learning can be seen from the readiness of e-learning that is planned because it is the most important part of distance education [2]. The level of community readiness in electronic learning needs to be considered by policymakers and stakeholders in designing and implementing efficient electronic learning programs [3]. In this study, the e-readiness data taken have similarities in measuring e-learning readiness, namely using the value scale on the Aydin & Tasci index model [2], [3], [4]. In this study, The author will conduct clustering analysis on e-learning readiness data based on the readiness level in the region and then visualize the readiness data by utilizing Geographic Information System (GIS) technology based on the region [5], [6].

The K-means method is the simplest and most common clustering technique [7], [8]. The K-means method is the most widely used clustering technique because it is made simpler. K-means can group a large amount of data with relatively fast and efficient computation time. The weakness of K-means is in determining the initial center of the clusters. This causes the results of the K-means cluster to be often stuck at the local optimal value [2]. In addition, the classification and clustering algorithms become problematic for data with high dimensions in the form of decreased classification accuracy and cluster quality. This problem can be overcome by reducing the dimensions. One of the methods of dimension reduction is principal component analysis (PCA) [10], [11], [12]. To reduce high-dimensional data into lower-dimensional data by minimizing the risk of losing very little information, PCA can be used. This study will use the PCA method as a data dimension reduction for cluster optimization on the K-means algorithm. Then evaluate clustering by using internal clustering validation metrics to select the optimal number of clusters between the K-means algorithm with K-means and

Received: 27 February 2026; Revised: 13 April 2026; Accepted: 29 April 2026.
*Corresponding author
[1]Eva Khudzaeva, UIN Syarif Hidayatullah Jakarta, Indonesia (e-mail: eva.khudzaeva@uinjkt.ac.id).
[2]Qurrotul Aini, UIN Syarif Hidayatullah Jakarta, Indonesia (e-mail: qurrotul.aini@uinjkt.ac.id).
[3]Evy Nurmiati, UIN Syarif Hidayatullah Jakarta, Indonesia (e-mail: evy.nurmiati@uinjkt.ac.id).
[4]Ismi Ana Sulasiyah, UIN Syarif Hidayatullah Jakarta, Indonesia (e-mail: ismi.ana17@mhs.uinjkt.ac.id).
[5]Ibrahim Shehu Usman, Department of Mathematics, Jigawa State College of Education, Nigeria (e-mail: babawoibrahim3@gmail.com).

PCA. The internal validation metric was chosen by the authors due to the absence of external information from the dataset used. The data grouping formed is divided based on the level of readiness in implementing e-learning. Then GIS is used to display information in a graphical form in the form of a map as an interface [13], [14], [15], [16].

Previous research on mapping using the K-means algorithm and PCA in the study [17] discussed the process of mapping lecturers based on performance in carrying out the education and teaching process by using one of the clustering techniques, namely K-means. In this study, PCA is combined with K-means. The PCA method is used to reduce the dimensions of the dataset before it is clustered using K-means. The cluster evaluation results show that the K-means algorithm, when preceded by reducing the dataset dimensions to 1 dimension using PCA, is more optimal with a DB Index value of 0.346, which is smaller compared to using K-means alone, and produces 4 clusters.

Based on the problems described previously, this study aims to group e-readiness data using clustering analysis the K-means method, then create a GIS-based map of the distribution of e-readiness clusters. The results of this study can be used as information on e-learning readiness, especially in Java, so that relevant decision makers can find out the potential of each region in implementing e-learning which can then be taken into consideration in implementing e-learning policies for every academic institution in Java.

II. Related Work

Previous studies have demonstrated [13] integrates spatial data and clustering methods using K-means, resulting in a web-based mapping system that groups crime levels and traffic problems. The clustering results show three sets of clusters, where four districts in cluster (C1) are classified as moderately vulnerable, four districts in cluster (C2) are classified as vulnerable, and four districts in cluster (C3) are classified as highly vulnerable to criminal acts. Traffic problem analysis also found that ten location points in cluster (C1) are classified as moderately vulnerable, eleven location points in cluster (C2) are classified as vulnerable, and seven location points in cluster (C3) are classified as highly vulnerable.

This research relies on GIS methods and multi-criteria decision-making, which helps reduce the number of comparisons in gathering expert opinions, simplifies the selection process, and improves assessment methods. It is used for selecting landfill sites (TPA) in the metropolitan city of Shiraz, Iran. Locations that meet disposal conditions are identified. These locations are then divided into six areas using the K-means clustering algorithm [5].

The researchers designed the potential level of each halal tourism destination to be recommended to the Ministry of Tourism of the Republic of Indonesia. The authors assessed the potential level based on the GMTI 2017 indicators and KM clustering. Simulations were carried out using 3, 5, and 7 clusters. The results showed that the more clusters there are, the better the accuracy and precision. In scenario 1, for three clusters, 46% of cases were in cluster number 2. The advantage of using KM on halal tourism destination data is that this method is easy to implement, and requires relatively faster and more flexible efforts to build the model [18].

Reference [19] implemented a GIS system mapping the distribution of poor families in Balikpapan City using K-means clustering on 44 attributes from 2020 social welfare data. The study demonstrated positive user acceptance, with 95.8% of 24 respondents strongly agreeing with the system's utility for identifying poor family locations by district.

A closely related study, [20] though slightly outside the 5-year window, combined K-means and Fuzzy C-means algorithms with GIS (using OpenStreetMap) to cluster and map agricultural lands in Southeast Minahasa by commodity type, providing visual information on land productivity and area distribution. Both studies demonstrate K-means' effectiveness for spatial clustering and GIS visualization in policy-relevant applications (poverty mapping and agricultural planning), though the provided sources contain limited Scopus-indexed publications explicitly combining these two methodologies in recent years.

Approach applies to K-means clustering within a GIS environment to classify traffic accident hotspots that were previously identified using Kernel Density Estimation (KDE). Hotspot areas are grouped according to accident characteristics and surrounding environmental factors, resulting in homogeneous clusters that represent distinct risk patterns. The results show that this method provides more comprehensive spatial insights than conventional accident analysis approaches and supports evidence-based decision-making in the development of traffic safety policies [7].

III. Research Method

This research includes quantitative research by collecting, analyzing, and displaying data in numerical form (numbers) rather than narrative. In Fig. 1, it shows the stages of research.

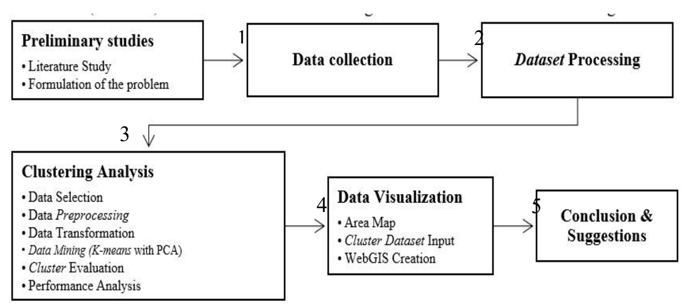

Fig 1. Research stages.

As shown in Fig. 1, regarding the research stages, the following is an explanation for each stage carried out:

1) In the first stage of the preliminary study, the authors conducted a literature study by collecting journal papers on e-readiness in Java and collecting related research journal papers to obtain the method to be used, then the authors formulated problems for research.
2) In the second stage of data collection, based on the problem formulation, the authors collect research in the form of journal papers, repositories, proceedings, conference papers regarding e-readiness in Indonesia.

Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 175–182
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50267

After collecting the data, the data is recapitulated in tabular form with a total of 41 locations of e-readiness data in Indonesia.

3) dataset processing, data filtering is carried out for the selected areas, namely: Java Island and filters for the selected e-readiness assessment model, namely: Aydin & Tasci index model. After filtering, the final total amount of research data used as a dataset is 27 locations and the research year is 2013–2021.
4) In the fourth stage of the clustering analysis process, referring to the data mining process, namely: data selection, data preprocessing, data transformation, data analysis PCA+K-means, and cluster evaluation using average within centroid distance, Davies Bouldin index (DBI), Cluster Density, and Sum Square as metrics to evaluate the results of the clustering algorithm. Based on the results of clustering validation and compiling a dataset that will be used in the GIS visualization stage.
5) In the fifth stage of GIS visualization, using a map of the island of Java with provincial and district/city boundaries, the author visualizes cluster data in the form of a web -based map (Web GIS) using the help of Quantum GIS (QGIS) tools.
6) In the last stage, the authors draw up conclusions and suggestions from the research that has been done.

IV. Result

A. Data Selection

At the data selection stage to be carried out, namely: selecting the attributes that will be used on the data. For the data mining process, the attributes used are academic level attributes, readiness scores, readiness categories, area areas, latitude (y) and longitude (x). Table 1 shows the results of the data selection that will be used for the next stage.

Table 1.
Results of Data Selection

Academic Level	Readiness Value	Readiness Category	Territory Area
College	3.78	Ready, requires implementation improvement	Banyumas Regency, Central Java
College	4.02	Ready, requires implementation improvement	Banyumas Regency, Central Java
College	3.98	Ready, requires implementation improvement	Banyumas Regency, Central Java
College	3.88	Ready, requires implementation improvement	Banyumas Regency, Central Java
College	3.46	Ready, requires implementation improvement	Bekasi Regency, West Java
College	4.04	Ready, requires implementation improvement	Cilacap Regency, Central Java
Senior high school	3.27	Not ready, requires some aspects of preparation	Klaten Regency, Central Java
College	2.81	Not ready, requires some aspects of preparation	Kuningan Regency, West Java
Vocational high school	3.45	Ready, requires implementation improvement	Sampang Regency, East Java
College	3.56	Ready, requires implementation improvement	Sleman Regency, DI Yogyakarta
College	3.7	Ready, requires implementation improvement	Sleman Regency, DI Yogyakarta
College	3.94	Ready, requires implementation improvement	Sleman Regency, DI Yogyakarta
College	3.42	Ready, requires implementation improvement	Bandung City, West Java
College	3.3	Not ready, requires some aspects of preparation	Depok City, West Java
College	3.53	Ready, requires implementation improvement	South Jakarta City, DKI Jakarta
Vocational high school	3.57	Ready, requires implementation improvement	South Jakarta City, DKI Jakarta
College	4.21	Ready, states that readiness is good	South Jakarta City, DKI Jakarta
Senior high school	4.31	Ready, states that readiness is good	South Jakarta City, DKI Jakarta
Vocational high school	3.42	Ready, requires implementation improvement	East Jakarta City, DKI Jakarta
College	3.38	Not Ready, requires some aspects of preparation	Madiun City, East Java
Elementary school	3.6	Ready, requires implementation improvement	Malang City, East Java
College	4.32	Ready, states that readiness is good	Malang City, East Java
College	4.05	Ready, requires implementation improvement	Semarang City, Central Java
College	3.68	Ready, requires implementation improvement	Surabaya City East Java
Junior high school	3.5	Ready, requires implementation improvement	Surakarta City, Central Java
College	3.6	Ready, requires implementation improvement	South Tangerang City, Banten
Junior high school	3.93	Ready, requires implementation improvement	Tangerang City, Banten

B. Preprocessing and Data Transformation

At this stage the data to be carried out will be transformed so that it is suitable for use in the data mining process. For example, the readiness category attribute is still nominal data type and must be converted to numeric data type (as seen in Fig. 2), so that the K-means and PCA algorithms can be run.

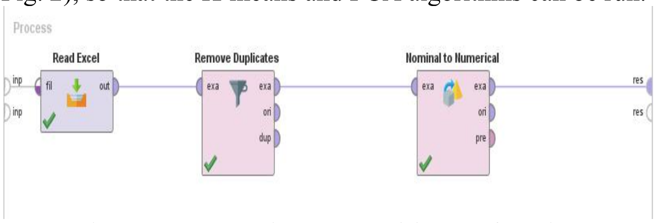

Fig 2. Data preprocessing process and data transformation.

Data cleaning and data transformation stages were successfully carried out and resulted in a dataset of 27 location data. The initialization of the readiness category column is shown in Table 2.

Table 2.
Initialization of Readiness Category

Readiness Category Initialization	Readiness Category
0	Ready, requires implementation improvement
1	Not Ready, requires some aspects of preparation
2	Ready, states that readiness is good

C. Clustering Analysis Stage

1) Application of the K-means and PCA algorithm

The implementation of the K-means and PCA algorithm in this study uses RapidMiner tools. The process of the PCA and K-means algorithms is shown in the Fig. 3.

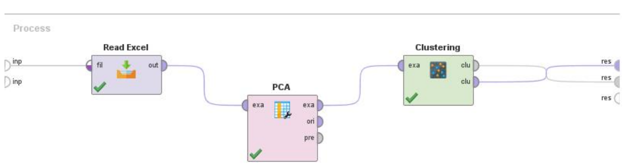

Fig 3. Process PCA and K-means modeling.

Figure 3 shows K-means and PCA modeling with RapidMiner. Based on Fig. 3, in the PCA operator the dimension reduction used is 1 dimension. The results of cluster segmentation can be seen Table 3.

Table 3.
Cluster Segmentation Results

K	Initial Area Area
2	Cluster _1(20 items): KbBms, KbBms, KbBms, KbBms, KbBks, KbClp, KbSpg, KbSmn, KbSmn, KbSmn, KBdg, KJktS, KJktS, KJktT, KSlg, KSmg, KSby, NGK Cluster _2(7 items): KbKln, KbKng, KDpk, KJktS, KJktS, KMad, KMlg
3	Cluster _1(20 items): KbBms, KbBms, KbBms, KbBms, KbBks, KbClp, KbSpg, KbSmn, KbSmn, KbSmn, KBdg, KJktS, KJktS, KJktT, KSlg, KSmg, KSby, NGK Cluster _2(4 items): KbKln, KbKng, Kdpk, KMad Cluster _3(3 items): KJktS, KJktS, KMlg
4	Cluster _1(4 items): KKln, KKng, KDpk, KMad Cluster _2(12 items): KbBks, KbSpg, KSmn, KBdg, KJktS, KJktS, KJktT, KMlg, KSby, KSkt, KTngS Cluster _3(3 items): KJktS, KJkts, KMlg Cluster _4(8 items): KbBms, KbBms, KbBms, KbBms, KbClp, KbSmn. KSmg, KTng
5	Cluster _1(3 items): KJktS, KJktS, KMlg Cluster _2(8 items): KbBms, KbBms, KbBms, KbBms, KbClp, KbSmn, KSmg, KTng Cluster _3(1 items): KbKng Cluster _4(12 items): KbBks, KbSpg, KbSmn, KbSmn, KBdg, KJktS, KJktS, KJktT, KMlg, KSby, KSkt, KTngS Cluster _5(3 items): KbKln, KDpk, KMad

2) Evaluation of K-means and PCA clustering results

To measure the quality capability of the cluster, clustering evaluation was carried out, which affects the number of *k* values. The evaluation of the number of clusters was conducted using 27 test data obtained from the results of k-means and PCA clustering in the previous stage. The dataset was then validated for clustering to determine the best number of clusters.

The evaluation was carried out using internal validation metrics, namely: Average within Centroid Distance (Av.D), Davies Bouldin Index (DBI), Cluster density, and Sum of squares.

The validation process is carried out with the help of RapidMiner tools. In the validation process of k-means clustering and PCA stage 1, the read excel operator and loop parameters operator are used. The read excel operator is used to read the dataset that will be tested. Then, the loop parameters operator is used to repeat the subprocess within this operator based on the predetermined combination of parameters. The parameters entered are k-means clustering with a minimum k value of 2 and a maximum k value of 5. The subprocess in the loop parameters is the clustering evaluation process using clustering validation metrics, namely: Average within centroid distance, DBI, cluster density, and sum square.

At the validation stage of k-means clustering and PCA stage 2, which is part of the loop parameters operator in stage 1, the process inside the loop parameters operator includes: PCA operator, k-means clustering operator, multiply operator, cluster distance performance operator (Av.D & DBI), data to similarity operator, cluster density performance operator, item distribution performance operator (sum square), and log operator. In the PCA operator, the selected dimensionality reduction value is 1 dimension, as in the k-means clustering and PCA modeling stage in the previous step. The graph of the evaluation results is shown in Fig. 4.

The graph in Fig. 4 shows the results of cluster validation with the K-means algorithm where each point of the k value shows the optimal value of each validation metric. For detailed results, see the performance comparison analysis section.

3) Performance analysis

To produce optimal performance analysis, compare 2 methods, namely K-means and K-means+PCA. Performance based on the number of clusters (*k* value) for each clustering algorithm is shown in Table 4.

Applied Information System and Management (AISM)
Volume 9, (1) 2026, p. 175–182
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50267

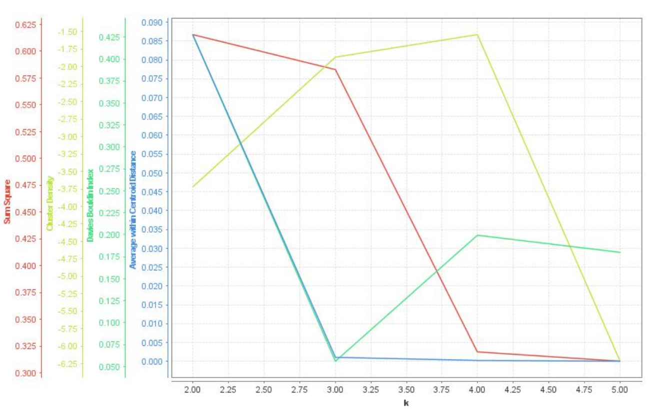

Fig 4. Line chart validation results average within centroid distance, Davies Bouldin index, cluster density and sum square on K-means and PCA algorithms.

Table 4.
Comparison of the Number of Clusters Based on the Validation Internal Clustering

Clustering Model	*k*	Clustering Validation Metrics			
		Av.D	DBI	Cluster density	Sum square
K-means	2	0.183	0.652	-6.665	0.616
	3	0.044	0.283	-4.377	0.583
	4	0.013	0.283	-2.305	0.320
	5	0.006	0.197	-7.261	0.311
K-means + PCA	2	0.087	0.428	-3.723	0.616
	3	0.001	0.057	-1.866	0.583
	4	0.000	0.200	-1,535	0.320
	5	0.000	0.180	-6,210	0.311

Based on Table 4, the average within centroid distance (Av.D) metric achieved its best value at 4 clusters for the k-means algorithm and 3 clusters for the k-means with PCA algorithm. To determine the best value based on the average within centroid distance metric: by selecting the smallest value among clusters. For the K-means algorithm, the best value obtained is 0.013, and for the PCA+K-means algorithm, the best value obtained is 0.001. After comparing these two values, the PCA+K-means value is the smallest, which is 0.001 with 3 clusters.

For DBI metric, the best value is obtained at 5 clusters, which is 0.197 for k-means, and at 3 clusters, which is 0.057 for PCA+K-means. The determination of the best value for the DBI metric is done by selecting the smallest value from each number of clusters. The smallest value between the two methods is 0.057 for the PCA+K-means method with 3 clusters. Then, for the cluster density metric, the best value is obtained at 5 clusters, which is -7.261 for K-means, and at 5 clusters, which is -6.210 for PCA+K-means. The selection of the best value for the cluster density metric is done by choosing the smallest value from each number of clusters. The smallest value from the two methods is -7.261 for the k-means method with 5 clusters.

After making a comparison between the two clustering methods, namely: k-means with k-means and PCA, it was found that the optimal value is achieved with the PCA+K-means method with 3 clusters. From these results, determining the best value for the average within centroid distance metric and the Davies Bouldin Index metric is influenced by the magnitude of each parameter item used. By using the PCA method as a dimensionality reduction, the attribute values can change the optimal number of clusters, which also affects the validation metrics Average within centroid distance and DBI that select the smallest value as the most optimal value, making the PCA+K-means cluster values smaller compared to using only K-means.

The addition of PCA to K-means also affects the density value of each cluster, where the use of PCA will increase the attribute values to approach zero (0) compared to the use of k-means. The sum square metric column shows the same value for each number of clusters, both for the k-means algorithm and PCA+K-means. This indicates that the performance of the sum square metric as a comparison between the two methods is not valid as a determinant of the optimal value between methods. After making comparisons between the two clustering methods, namely: K-means with K-means and PCA, the optimal value was found to be in the PCA+K-means method with 3 clusters. With the selection of cluster results using the K-means and PCA algorithms, the author then defines the readiness level of each cluster. With the selection of cluster results using the K-means algorithm and PCA, the author then defines the readiness level of each cluster.

Table 5
Tabel Centroid Algorithm K-means+PCA

Attribute	Cluster_1	Cluster_2	Cluster_3
pc_1	-0.364	0.641	1.703
Readiness category	0	1	2
Readiness score	3.7	3.38	4.21

To define the readiness level, the author refers to the centroid table, namely: Table 5, which shows that the row pc_1 is the centroid value resulting from a 1-dimensional reduction on the readiness value attribute and readiness category. In this case, the author compares the centroid value of pc_1 with the readiness value attribute and readiness category from the preprocessed dataset used in the PCA+k-means stage. The author then categorizes the e-learning readiness level in Java Island as follows: **cluster 1** represents a collection of e-readiness data with a ready level with requires implementation improvement, **cluster 2** with a not ready level, and **cluster 3** with a ready level (Fig. 5).

By combining the cluster attributes from the K-means and PCA algorithms into a dataset that will be used to visualize GIS-based E-readiness data. The clustering data used in the GIS stage is shown in Table 6.

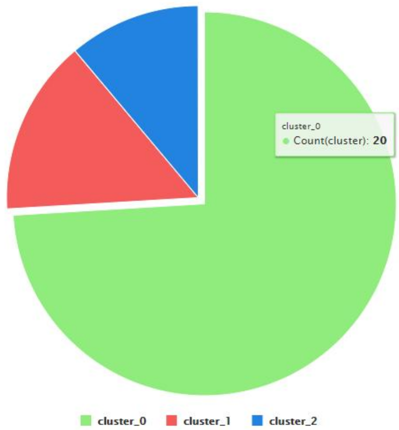

Fig 5. Visualization of E-learning readiness distribution data in the form of pie chart.

Table 6.
E-Readiness Clustering Data in Java Island

Year	Academic level	Cluster	Territory Area	*y*	*x*
2015	College	cluster_1	Bandung City, West Java	-6.9730	107.6316
2016	College	cluster_1	Sleman Regency, DI Yogyakarta	-7.7613	110.4246
2017	College	cluster_1	South Jakarta City, DKI Jakarta	-6.3107	106.7935
2017	College	cluster_1	South Jakarta City, DKI Jakarta	-6.2838	106.7947
2019	College	cluster_1	East Jakarta City, DKI Jakarta	-6.1943	106.8866
2019	College	cluster_1	Surakarta City, Central Java	-7.5581	110.8090
2020	Senior high school	cluster_1	Bekasi Regency, West Java	-6.3239	107.1692
2020	College	cluster_1	Sampang Regency, East Java	-7.0594	113.1702
2020	Vocational high school	cluster_1	Sleman Regency, DI Yogyakarta	-7.7748	110.3862
2020	College	cluster_1	Sleman Regency, DI Yogyakarta	-7.7804	110.4141
2020	College	cluster_1	Malang City, East Java	-8.0011	112.6141
2020	College	cluster_1	South Tangerang City, Banten	-6.2668	106.7323
2021	College	cluster_1	Banyumas Regency, Central Java	-7.4394	109.2662
2021	College	cluster_1	Banyumas Regency, Central Java	-7.4349	109.2493
2021	College	cluster_1	Banyumas Regency, Central Java	-7.4107	109.2316
2021	Vocational high school	cluster_1	Banyumas Regency, Central Java	-7.4042	109.2463
2021	College	cluster_1	Cilacap Regency, Central Java	-7.7174	109.0199
2021	Senior high school	cluster_1	Semarang City, Central Java	-7.0528	110.4344
2021	Vocational high school	cluster_1	City of Surabaya, East Java	-7.3045	112.7338
2021	College	cluster_1	Tangerang City, Banten	-6.2054	106.6774
2016	Elementary school	cluster_2	Depok City, West Java	-6.3157	106.7945
2018	College	cluster_2	Madiun City, East Java	-7.6341	111.5420
2019	College	cluster_2	Klaten Regency, Central Java	-7.6861	110.6132
2020	College	cluster_2	Kuningan Regency, West Java	-6.9749	108,5004
2013	Junior high school	cluster_3	South Jakarta City, DKI Jakarta	-6.2182	106.8296
2020	College	cluster_3	South Jakarta City, DKI Jakarta	-6.3359	106.8300
2020	Junior high school	cluster_3	Malang City, East Java	-7.9517	112.6074

D. *GIS Visualization Stage*

In the GIS visualization stage where the clustering results that have been formed are then visualized in the form of a GIS-based map. The map layer is shown in Fig. 7.

Fig 7. Java island map layer with E-readiness level cluster points.

Based on Fig. 7, the map is divided by Province on the island of Java which is divided into: aqua color for Banten, purple for DKI Jakarta, green for West Java, orange for Central Java, yellow color for DI Yogyakarta, and magenta color for East Java. This map also contains Regency/City information for each Province. After obtaining the Provincial and Regency/City layers on the island of Java, in the next stage input will be carried out clustering data that has been obtained at the clustering analysis stage. Each cluster is indicated by a different color. In cluster _1 use green which indicates the level of readiness, in cluster _2 using red which indicates the level of less ready, and in cluster _3 using blue color which states the very ready level. After entering the cluster dataset, the next step is to create a Web GIS.

Volume 9, (1) 2026, p. 175–182
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v9i1.50267

The final stage of GIS visualization is the creation of Web GIS as the result of this research. The result based on Web GIS aims to make it easier for readers to access the results of this research because Web GIS can be accessed using a web browser which is currently owned by all internet users. In addition to facilitating data access, readers will find it easier to understand the level of e-learning readiness due to easy navigation. Web GIS creation using the web2gis plugin available in QGIS. The web browser used by the author to display Web GIS is Google Chrome. The display of Web GIS in Google Chrome browser can be seen in Fig. 8.

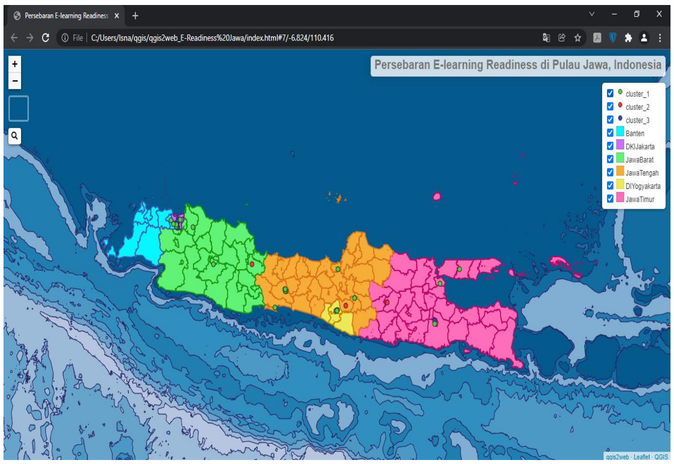

Fig 8. WebGIS visualization of E-learning readiness distribution map in Java island.

V. CONCLUSION

This research has succeeded in conducting clustering analysis on the dataset e-learning readiness in Java, totaling 27 items with 2 parameters used based on the level of readiness. The number of clusters tested were: 2, 3, 4, and 5 clusters with the K-means method. Cluster was determined based on the internal validation metric clustering, namely: 3 clusters. The number of 3 clusters can be defined: cluster 1 (20 locations) for ready level, cluster 2 (4 locations) for less ready level, and cluster 3 (3 locations) for very ready level.

The resulting GIS visualization is based on Web GIS with the distribution based on the Province in Java Island. For Banten Province there are 2 clusters of ready level, for DKI Jakarta Province there are 3 clusters of ready level and 2 clusters of very ready level, for West Java Province there are 2 clusters of less ready level and 2 clusters of ready level, for Central Java Province there is 1 cluster of less level ready and 7 clusters ready level, for DI Yogyakarta Province there are 3 ready level clusters, and for East Java Province there is 1 less ready level cluster, 2 ready level clusters, and 1 very ready level cluster.

Internal clustering validation metrics used are Average within Centroid Distance, Davies Bouldin Index, Cluster Density, and Sum of Square. Compared to similar research, this study has the advantage of adding a sum of square metric that has not been combined with other internal validation metrics. The metric value Average within Centroid Distance is: 0.001 with a number of clusters of 3, for the Davies Bouldin Index (DBI) metric the number of clusters 3 is: 0.057 and for the Cluster density metric the number of clusters 5 is: -6.210

This study is expected to provide practical impacts on the field of education, particularly e-learning, as information on the potential of each region in the implementation of e-learning so that it can be used as a consideration for relevant decision-makers in applying e-learning policies for each academic institution in Java Island.

This study has several limitations, namely that the e-learning readiness dataset used only comes from research that has been published freely, allowing the dataset to be supplemented with direct institutional data. Future studies should expand the diversity of samples and use mixed-method data collection to strengthen the findings, and the geographical coverage should also be expanded, not only on the island of Java. For further optimization, it is possible to combine the k-means method with other optimization algorithms, such as genetic algorithms, Particle Swarm Optimization (PSO), Support Vector Machine (SVM), Hierarchical Clustering, DBSCAN, or Gaussian Mixture Models, and so on.

REFERENCES

[1] A. H. H. Rohayani, K. Kurniabudi, and S. Sharipuddin, “A literature review: Readiness factors to measuring e-learning readiness in higher education,” *Procedia Comput. Sci.*, vol. 59, pp. 230–234, 2015, doi: 10.1016/j.procs.2015.07.564.

[2] A. A. Mosa, M. N. B. Mahrin, and R. Ibrrahim, “Technological aspects of E-learning readiness in higher education: A review of the literature,” *Computer and Information Science*, vol. 9, no. 1, Art. no. 113, 2016, doi: 10.5539/cis.v9n1p113.

[3] P. Kaur, M. Singh, and G. Singh, “Classification and prediction based data mining algorithms to predict slow learners in education sector,” *Procedia - Procedia Comput. Sci.*, vol. 57, pp. 500–508, 2015, doi: 10.1016/j.procs.2015.07.372.

[4] C. Aydin and D. Tasci, “Measuring readiness for e-learning: reflections from an emerging country,” *Journal of Educational Technology & Society*, vol. 8, no. 4, pp. 244–257, Jan. 2005. [Online]. Available: http://www.jstor.org/stable/jeductechsoci.8.4.244

[5] M. Eghtesadifard, P. Afkhami, and A. Bazyar, “An integrated approach to the selection of municipal solid waste landfills through GIS, K-Means and multi-criteria decision analysis,” *Environmental Research*, vol. 185, Art. no. 109348, Mar. 2020, doi: 10.1016/j.envres.2020.109348.

[6] S. Handoko, E. Sediono, and S. Suhartono, “Sistem informasi geografis berbasis web untuk pemetaan sebaran alumni menggunakan metode K-means,” *Jurnal Sistem Informasi Bisnis*, vol. 1, no. 2, pp. 80–85, Jul. 2011, doi: 10.14710/interaksi.%v.%i.145-152.

[7] T. K. Anderson, “Kernel density estimation and K-means clustering to profile road accident hotspots,” Accident Analysis & Prevention, vol. 41, no. 3, pp. 359–364, Jan. 2009, doi: 10.1016/j.aap.2008.12.014.

[8] M. Sadeghi, P. Casey, E. J. M. Carranza, and E. P. Lynch, “Principal components analysis and K-means clustering of till geochemical data: Mapping and targeting of prospective areas for lithium exploration in Västernorrland Region, Sweden,” *Ore Geology Reviews*, vol. 167, Art. no. 106002, Mar. 2024, doi: 10.1016/j.oregeorev.2024.106002..

[9] D. Suyanto, *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika Bandung, 2017.

[10] A. Jamal, A. Handayani, A. A. Septiandri, E. Ripmiatin, and Y. Effendi,

"Dimensionality reduction using PCA and K-means clustering for breast cancer prediction," *Lontar Komputer Jurnal Ilmiah Teknologi Informasi*, vol 9, no.3, pp. 192–201, Dec. 2018, doi: 10.24843/lkjiti.2018.v09.i03.p08.

[11] N. I. Ali, S. Samsuri, I. A. Brohi, A. B. Soomro, S. Soomro, and A. Shah, "Preliminary study on factors affecting e-commerce success: A modified Delone and McLean model," *Proc. - Int. Conf. Inf. Commun. Technol. Muslim World 2018, ICT4M 2018*, pp. 120–125, 2018, doi: 10.1109/ICT4M.2018.00031.

[12] G. P. Rédei, "PCA (principal component analysis)," *Encycl. Genet. Genomics, Proteomics Informatics*, vol. 6, no. 4, pp. 1457–1457, 2008, doi: 10.1007/978-1-4020-6754-9_12426.

[13] A. Pratama, M. D. Irawan, and S. D. Andriana, "Implementation of K-means clustering in recognizing crime hotspots and traffic issues through GIS," *Journal of Computer Networks Architecture and High Performance Computing*, vol. 6, no. 2, pp. 771–782, Apr. 2024, doi: 10.47709/cnahpc.v6i2.3771.

[14] S. Syaifuddin, N. A. M. Puad, and S. Rusdian, "The role of geographic information systems in marketing strategy: Improving the efficiency of store locations and consumer targeting based on Alfamart locations in Bandung City," *Applied Information System and Management (AISM)*, vol. 8, no. 2, pp. 213–220, Oct. 2025, doi: 10.15408/aism.v8i2.45522.

[15] D. Festa *et al.*, "Unsupervised detection of InSAR time series patterns based on PCA and K-means clustering," *International Journal of Applied Earth Observation and Geoinformation,* vol. 118, Art. no. 103276, Mar. 2023, doi: 10.1016/j.jag.2023.103276..

[16] C. N. S. Kumar, "Spatial data mining using cluster analysis," *International Journal of Computer Science and Information Technology,* vol. 4, no. 4, pp. 71–77, Aug. 2012, doi: 10.5121/ijcsit.2012.4407.

[17] A. Izzuddin, "Optimasi cluster pada algoritma K-means dengan reduksi dimensi dataset menggunakan principal component analysis untuk pemetaan kinerja dosen," *Energy*, vol. 5, no. 2, pp. 41–46, 2015.

[18] Q. Aini and E. Khudzaeva, "Potential halal tourism destinations with applying K-means clustering," *International Journal of Intelligent Systems and Applications,* vol. 11, no. 9, pp. 9–17, Sep. 2019, doi: 10.5815/ijisa.2019.09.02.

[19] A. Rivaldi, J. D. Irawan, and A. Faisol, "A geographic information System for mapping the distribution of poor families using the K-Means method," (Indonesian), *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 4, pp. 2436–2442, Jan. 2024, doi: 10.36040/jati.v7i4.7541.

[20] J. Tamaela, E. Sediyono, and A. Setiawan, "Cluster analysis using the fuzzy C-means and K-means algorithms for the clustering and mapping of agricultural land in Southeast Minahasa," (Indonesian), *Jurnal Buana Informatika*, vol. 8, no. 3, pp. 151–160, Oct. 2017, doi: 10.24002/jbi.v8i3.1317.