Advanced Analytical Of Use Of Electronics Government Using K-Means Algorithm

Ibnu Teguh Ghifary, Deden Witarsyah, Rachmadita Andreswari, Ahmad Musnansyah

Abstract: E-Government is a public service carried out by all government agencies that are coordinated with each other optimally using telematics technology. In order for the implementation of e-government to be carried out properly, it is necessary to consider technical and non-technical factors that can influence success. In general, non-technical factors are more dominant than technical factors, therefore an in-depth understanding of non-technical factors is needed when designing and implementing e-government. The implementation of e-government that is associated with efforts to meet the needs of all sectors of activity both in government and society requires patrons from leaders who can provide examples and shared commitment. For this reason, an analysis of interest in the use of e-government using the K-means algorithm on RapidMiner tools that uses supporting variables is a survey question of service factors, namely performance expectancy, effort expectancy, and social influence to be studied. The output of this study is the interest in using e-government to the UTAUT factor which is more dominant in the community. Based on the results of the study, K-means has an accuracy rate of analysis that can reach 91%.

Index Terms: Electronic Government, Performance expectancy, Effort expectancy, Social influence, K-Means, Clustering, Rapid Miner.

1. INTRODUCTION

The use of information technology is developing rapidly and now almost all can use technology. The development of this technology is no exception used by state governments to assist in carrying out its obligations. Electronic-Government (E-gov) is an application of Information Technology in government which aims to make the work process in government simple, more accurate, responsive and form a transparent government. The using of e-government can enable the wider community to access all government information and services through a website that is managed by the government [1]. The operational advantages and implementation of e-government for government and public sector companies such as reducing paper, providing services available to customers, reducing response times and reducing errors in e-government services for the community[2]. The implementation of e-government in Indonesia has obstacles or obstacles that make e-government in Indonesia can experience failure in its implementation. This can be detrimental to the government itself and the lack of public interest in the use of e-government, because in addition to incurring large costs and quite a long time but can not provide services in e-government to the fullest [3]. From 2015 to 2018, internet usage in the community amounted to 72.7 million to 132.7 million with a percentage of 82%. Then in the survey results of the Indonesian Internet Service Providers Association (APJI), the percentage of Indonesian people is still low in accessing e-government services, which is no more than an average of 15%. This can cause a lack of public awareness in accessing e-government services and the internet is only used for lifestyle purposes only (Barsei, 2018). In the use of e-governement, there are several models of factors that influence including performance expectancy (measuring one's level of trust), effort expectancy (individual efforts to use e-government), and social influence (one's belief in others about the use of e-government). These three factor models are measurements to find out someone's interest in using an information technology system and subsequent user behavior [4].

The K-Means algorithm was chosen because it is one of the non-hierarchical methods that partition existing data into one or more clusters and also has a high degree of accuracy in object size, so this algorithm is relatively more measurable and efficient. By using the clustering method, it can search and group data that has similar characteristics between the data one with other data that has been obtained. Several methods have been used in e-government analysis, namely in determining the status of EDGI (E-Government Development Index). However, the determination of EDGI is less accurate because it must be based on knowledge and processing of the amount of data [5]. In the study [7] to initialize the initial cluster center in the K-Means Algorithm, it provides ease of implementation and can also handle large amounts of data. Based on the aforementioned matters, an analysis which raises the issue will be carried out by applying clustering methods that focus on k-means algorithm on survey data conducted. towards e-government. Researchers hope that the results of clustering can be used to determine interest in the use of e-government in society so that it can have a major influence on its use in Indonesia.

- *Ibnu Teguh Ghifary is currently pursuing bachelor degree program in Information System Department in Telkom University, Bandung, Indonesia. E-mail: ghifaryibnu @gmail.com*
- *Deden Witarsyah is currently pursuing lecturer in Information System Department in Telkom University, Bandung, Indonesia. E-mail: dedenw@telkomuniversity.ac.id*
- *Rachmadita Andrewari is currently pursuing lecturer in Information System Department in Telkom University, Bandung, Indonesia. E-mail: andreswari@telkomuniversity.ac.id*
- *Ahmad Musnansyah is currently pursuing lecturer in Information System Department in Telkom University, Bandung, Indonesia. E-mail: ahmadanc@telkomuniversity.ac.id*

2 LITERATURE REVIEW

2.1 E-Government

According to witarysah [6], E-Government is the use of information technology by government agencies such as wide area networks (WAN) internet, moble competing that can be used to build relationships with the public, business world and other government agencies. Meanwhile, according to The World Bank Group is an effort to use information and communication technology to improve efficiency and effectiveness [7]. Based on the results of a 2012 survey by the United Nations, Malaysia ranked second after Singapore as the best country in developing e-government in Southeast

Asia. Whereas Indonesia ranks 7th out of 11 countries implementing e-government in 2012 [8] (see figure 1).

Country	E-Government 2012	Rank 2012	Rank 2010	Rank Change
Singapore	0.8474	10	11	+1
Malaysia	0.6703	40	32	-8
Brunei Darussalam	0.6250	54	68	+14
Viet Nam	0.5217	83	90	+7
Philippines	0.5130	88	78	-10
Thailand	0.5093	92	76	-16
Indonesia	0.4949	97	109	+12
Lao People's Democratic Republic	0.2935	153	151	-2
Cambodia	0.2902	155	140	-15
Myanmar	0.2703	160	141	-19
Timor-Leste	0.2365	170	162	-8

***Figure 1**. World E-government Rangking*

2.2 Data Mining

Data mining is about solving problems by analyzing data in the database and is often also referred to as the process of finding patterns in data where the process must be automatic or semi-automatic and the patterns found must be meaningful [9]. Simply put, data mining can also be interpreted as the process of extracting or digging into existing knowledge in a set of data. The information and knowledge obtained can be used in many fields, such as business management, education, health and so on.

The steps of data mining can be described as follows:

1. Data cleaning, eliminating noise and inconsistent or irrelevant data
2. Data integration, merging data from various databases into one new database
3. Data Selection, retrieve data relevant to the analysis task from the database
4. Data transformation, transform or merge data into forms suitable for processing in data mining
5. Mining process, a main process when the method is applied to find valuable and hidden knowledge from data.
6. Knowledge presentation, a visualization and presentation of knowledge about the method used to obtain the knowledge.

2.3 K-Means Algortihm

K-Means is a non-hierarchical cluster analysis method that attempts to partition objects into one or more cluster groups of objects based on their characteristics, so objects that have the same characteristics are grouped in the same cluster and objects that have different characteristics are grouped into another cluster. The K-means formula can be explained as follows:

$$D(i,j) = \sqrt{(X_{1_i} - X_{1_j})^2 + (X_{2_i} - X_{2_j})^2 + \ldots + (X_{k_i} - X_{k_j})^2}$$

where:
D (i, j) = Distance of data to i to cluster center j
Xki = Data to i in attribute data to j in attribute to k
Xkj = Center point to j in attribute to k

K-Means which are implemented using the Euclidean Distance calculation algorithm provide the best results and provide accurate results [10, 14]. Correlation is a value that shows the closeness of the relationship between two variables. Two variables are called positively correlated when the two variables experience a similar increase. Two variables are called negatively correlated when the two opposing variables (one variable is impaired). Two variables are called uncorrelated when the two variables are not linearly related to each other. The following is the formula for calculating correlations in the sample data as follows:

$$kor(x,y) = r_{x_y} = \frac{\sum_{i=1}^{n}(x_i - x)(y_i - y)}{\sqrt{\sum_{i=1}^{n}(x_i - x)^2 \sum_{i=1}^{n}(y_i - y)^2}}$$

Whereas for population data are:

$$kor(x,y) = \rho_{x_y} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

where:
Σ = to calculate the sum
$(X_i - \bar{x})$ = value of x_i minus the average, $\bar{x}$
$(y_i - \bar{y})$ = value of y_i minus the average, $\bar{y}$
cov (x, y) = correlation of x and y
$\sigma_x \sigma_y$ = standard deviation of x and y

3. RESEARCH METHODOLOGY

The research methodolgy in Figure 2 will be explained as follows [11-13]:

1. Collection Phase

At the data collection stage is the initial and most important stage that must be done. Before moving to the next stage, the data that has been collected needs to be understood in advance to find out what information can be obtained from the data.

2. Selection Stage

In the selection of data, choose the most influential variables from the data so that the results can be more accurate. Selecting data can minimize the scope so that research will be focused and targeted.

3. Changing Stage

At the stage of modification that is changing the form of data to fit the algorithm used. This stage must be done because there are several algorithms that do not perform well when the data values do not meet the requirements.

4. Data Mining Stage

In the data mining phase, it aims to obtain information on e-government adoption from survey data that has been conducted. This stage starts from the task of data mining which is clustering. Data that performs the cleaning process will be clustered based on the question data attributes contained in each e-government service. This prediction will be calculated using the k-means algorithm. Data that has been

purged will be processed to obtain the dominant adoption level output in the three e-government services studied.

5. Evaluation Phase

After the data mining process is complete and the probability of predicting the level of adoption of e-government services, the data will be evaluated and processed by reading the results of the k-means algorithm.

***Figure 2**. Systematic Research Methodology*

4 RESULT

The data used in this study are data taken from the previous research [1]. The sample data in Table 1 consists of a 7 column test consisting of a peformance expectancy question.

***Table 1** Dataset Performance Expectancy*

Profession	PE1	PE2	PE3	PE4	PE5	PE6	PE7
Student	3	3	2	4	4	2	3
Exe. Manager	4	5	5	5	5	5	3
Employee	3	3	3	5	3	3	3
Lawyer	4	4	4	4	4	3	3
Teacher/Lecturer	4	3	4	4	4	2	1
Medical Doctor	5	5	5	5	4	3	2
Secretary	3	4	3	4	4	2	2

There are five names that are used, they are strongly disagree, disagree, neutral, agree, and strongly agree. To determine the accuracy of the K-Means algorithm, there needs to be a comparison between the actual planting and naming of the predicted results. A total of 237 people who have filled out the correspondence will be selected in accordance with the analysis and adjusted to the naming given. Data that has been divided in accordance with the naming and then will do the clustering process using the K-Means algorithm that has been tested for accuracy {15-17] (see table 2). Here are the results of clustering using the K-Means algorithm for service expectancy service factors and also testing the accuracy of the other two service factors, namely effort expectancy and social influence.

***Table 2** Accuracy Level of Each Cluster*

	Accuracy Level				
	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Peformance Expectancy	51%	53%	43%	79%	11%
Effort Expectancy	52%	29%	91%	29%	36 %
Social Influence	37%	63%	85%	11%	41 %

In table 2, it can be seen that the accuracy of each service is different for each service. In the table there is an accuracy value from each cluster. In the Peformance Expectancy service with 237 data in which 44 data are stated strongly agree with 79% accuracy contained in cluster 3, 91 data are stated agree with 53% accuracy contained in cluster 1, 90 data are stated to be normal with 51% accuracy contained in the cluster 0, 6 data were stated to disagree with the accuracy of 43% contained in cluster 2, and 6 data were stated strongly disagree contained in cluster 4. Furthermore, in the effort expectancy service with 237 data in which 28 data were stated strongly agree with 91% accuracy contained in cluster 2, 90 data were stated agreed with 52% accuracy contained in cluster 0, 52 data were stated to be normal with an accuracy of 36% contained in cluster 4, 37 data were declared not in agreement with an accuracy of 29% contained in cluster 3, and 30 the data is stated strongly disagree with the accuracy of 29% contained in cluster 1. In social influence service with 237 data, where 83 data are stated that agree with 85% accuracy contained in cluster 2, 15 data are stated agree with 63% accuracy contained in cluster 1, 15 data are stated as normal with 41% accuracy contained in the cluster 4, 71 data were stated to disagree with 37% accuracy contained in cluster 0, and 53 data were stated strongly disagree with 11% accuracy contained in cluster 3. Clusters with the highest accuracy were positive statements from the questionnaire given, while clusters with accuracy the lowest is a negative statement. From the above accuracy will be taken the highest cluster of each service and then will be compared as in Table 3 as follows.

***Tabel 3**. Accuracy Comparison Results*

Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	
6,3%	0%	16,9%	4,9%	10,1%	
Centroid Distance					8,2%

In Table 3, it can be seen that the comparison between the three services has a different level of accuracy. In cluster 2 it has a high accuracy rate of 16.9% while in cluster 1 it has a very low accuracy of 0%. The distance between each is determined by a centroid distance of 8.2%. After clustering using the K-Means algorithm, we can find out how e-government is adopted in government. This can be seen from Figure 3.

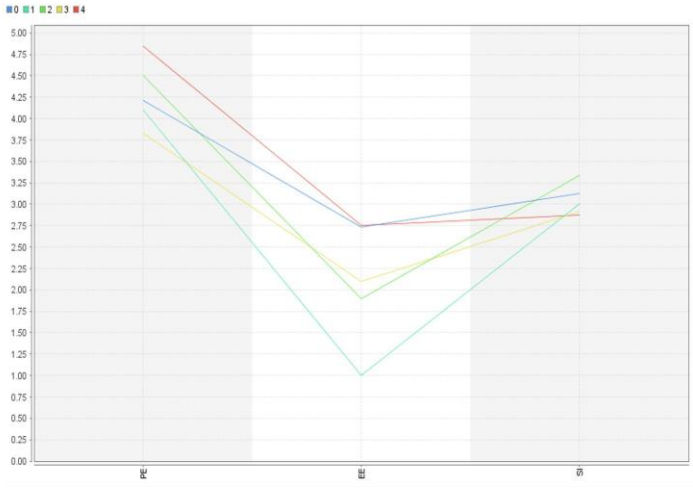

***Figure 3**. Line plot of three services*

Figure 3 is a line plot of three services that have been compared, namely PE (peformance expectancy), EE (effort expectancy), and SI (social influence). In the picture there are also colors and numbers that indicate the naming of the clusters produced using the K-Means algorithm. The diagram shows cluster 4 strongly agrees, cluster 0 agrees, cluster 2 is normal, cluster 3 disagrees, and cluster 1 strongly disagrees

with the adoption of e-government. During this period, the performance expectancy service experienced excellence between the other two services.

4 CONCLUSION

Based on research that has been done, it can be concluded thatClustering using the K-Means method in e-government in government is done by taking data from the research of Mr. Deden Witarsyah, then selecting three services namely peformance expectancy, effort expectancy and social influence which then collect data to clean up the data so that it can be processed in the clustering stage so as to produce cluster values from 0 to 4 that will be given in accordance with the naming. In the three e-government services in government namely the service expectancy service factor 79% with the amount of data 1,659 contained in the cluster 3. For the service expectancy service factor 91% with the amount of data 1,422 and the social service factor influence 85% with a total of 948 data contained in cluster 2 each. The experiments conducted on peformance expectancy, effort expectancy, and social influence services the highest accuracy is taken from each cluster. Then the new cluster will produce an average accuracy value of 8.2%. After that, in the plot line diagram, the factors of peformance expectancy services are adopted more by Indonesian citizens compared to the other two services. The striking correspondent question on "I find that dealing with government employees directly is more efficient than dealing with them using the internet" which has a negative type of question with a cluster value of 4.66 (Disagree). Next, It is hoped that for further research, data processing that has attributes that describe areas of society that are more specific in filling in correspondent questions will make it easier in the clustering process.

REFERENCES

[1] Jacob, D. W., Fudzee, M. F. M., Salamat, M. A., & Herawan, T. (2019). A review of the generic end-user adoption of e-government services. International Review of Administrative Sciences, 85(4), 799–818. https://doi.org/10.1177/0020852319861895.

[2] Indra Sarjono Sipahutar, S. s. (2017). Jurnal Riset Akuntansi dan Keuangan. Faktor-Faktor Penentu Implementasi E-Goverment Pemerintah Daerah Di , 18-21.

[3] https://www.kompasiana.com/royandi/5c0a2747aeebe16fe8283cc2/dampak-e-government-dalam-kemajuan-di-bidang-pemerintahan

[4] Venkatesh V, Morris M, Davis G, et al. (2003) User acceptance of information technology: Toward a unified view. MIS Quarterly 27(3): 425–478.

[5] Neesha Jothi, N. A. (2015). Journal School of Computer Sciences Universiti Sains Malaysia. Data Mining in Healthcare.

[6] Witarsyah, D., Fudzee, M. F. M., Salamat, M. A., Yanto, I. T. R., & Abawajy, J. (2020). Soft Set Theory Based Decision Support System for Mining Electronic Government Dataset. International Journal of Data Warehousing and Mining, 16(1), 39–62. doi:10.4018/ijdwm.2020010103

[7] https://www.worldbank.org/en/topic/digitaldevelopment/brief/digital-government-for-development

[8] United Nation e-Government Survey 2014

[9] Yudi Prastyo, P. E. (t.thn.). Jurnal Universitas Ibn Khaldun Bogor. Pembagian Tingkat Kecanduan Game Online Menggunakan K-Means Clustering Serta Korelasinya Terhadap Prestasi Akademik.

[10] Ahmar, A. S. (2018). Journal of Physics: Conference Series. Using K-Means Clustering to Cluster Provinces in Indonesia.

[11] Becker, J., Niehaves, B., & Janiesch, C. (2010). Socio-Technical Perspectives on Design Science in IS Research. Information Systems and eBusiness Management, Vol.9, issue 1, 109-131.

[12] Hevner, A. R., Ram, S., March, S. T., & Park, J. (2004). Design Science in Information Systems Research. MIS Quarterly Vol. 28 No. 1, 75-105.

[13] Hevner, A., & Chatterjee, S. (2010). Design Research in Information System : Theory and Practice. New York: Springer.

[14] Rishikesh Suryawanshi, S. P. (2016). International Journal of Computer Applications. A Novel Approach for Data Clustering using Improved K-Means Algorithm, 13-17.

[15] Shenna Angra, S. A. (2016). Journal Chitkara University India. Analysis of Student's Data Using RapidMiner, 112-114.

[16] Tapas Kanugo, N. S. (2002). Journal IEEE Transactions On Pattern Analysis and Machine Intelligence. An Efficient K-Means Clustering Algortihm: Analysis and Implementation.

[17] Widowati, D. P. (2018, Agustus 23). Survei PBB 2018: Peringkat E-Government Indonesia.