NEWTON: Networking and Information Technology
Vol. 2 No. 2 October 2022, Page. 89 - 98
E-ISSN: 2797-0728

Implementation of The K-Mean Algorithm for Data Clustering in Alumni Data Information Systems

Agus Sifaunnajah[1*], Muhyiddin Zainul Arifin[2] Habib Qoirul Tabi'in[3]
[1,2,3] Information System, Universitas KH. A. Wahab Hasbullah
Email: agus.syifa85@gmail.com

ABSTRACT
Graduates are the most important parameter of an organization, alumni are also active for a formal or informal educational organization or institution. In this era of big data, the correct and fast management of alumni data can bring significant benefits to the institution concerned. Algorithm K-means Clustering is one of the data collection algorithms that ranks data quite well. Using a research method, the Research and Development (R&D) department will try, as far as possible, to produce works that can be used to test real, not just hypotheses or theories.

Keywords: *K-means; Clustering; R&D; Alumni*

INTRODUCTION
Alumni are the most important parameter in an organization, Alumni are also an asset for an organization or formal and non-formal educational institutions, alumni can also be a reflection of quality for the institution, especially in Islamic boarding schools that graduate their students every year. In this era of big data, managing alumni data properly and quickly can provide significant benefits for the institution concerned. With a large amount of alumni data, many problems occur, especially in the problem of finding alumni data in a fast and accurate way (Destari, 2015).

Clustering or clustering is a method in data mining that is used to analyze data to find problems in grouping data or rather partitioning the dataset into subsets (Nugroho *et al*., 2020). In the clustering method, the targets are cases of distribution (objects, people, events, and others) into a group, so that the degree of connectedness between members of the same cluster is strong and weak between members of different clusters (Ali, 2019).

The K-Mean algorithm is one of the data mining algorithms that are quite good at clustering data. The K-mean algorithm in some studies managed to do a good job of clustering data. In this study, the K-mean Algorithm will be applied to the alumni data information system to conduct data clusters based on several criteria set by the foundation management based on the request (Dhuhita, 2015).

Based on observations and interviews with the management of the Jombang Bahrul Ulum Islamic Boarding School Foundation, alumni data collection was carried out computerized, namely this activity was carried out by a team from the Jombang Bahrul Ulum Islamic Boarding School Foundation by the manual method through questionnaires. This causes the length of the process of collecting job data and alumni income. In addition, sometimes the available information is not up to date so the information becomes less accurate for the success of alumni grouping based on poverty indicators(Putrawansyah & Dewi, 2021).

Based on the purpose of the journal above, the benefits that can be provided are being able to help the management of the Bahrul Ulum Islamic Boarding School Foundation in Jombang rice ponds in finding alumni data quickly and concisely and knowing between successful and unsuccessful alumni based on poverty indicators.

The research method used is research and development (R&D). The data used is data on alumni of the Bahrul Ulum Islamic Boarding School Tambakberas Jombang. Data analysis is qualitative by applying the system development life cycle. By using this method, the development will try as much as possible to produce a work that can be used in real terms, not only testing certain hypotheses or theories. This research and development is carried out and prioritized on the application of the K-mean algorithm for data management of alumni of the bahrul ulum boarding school in Jombang rice ponds.

METHOD

Process flow in Software

Flowchart is a graphical representation of an algorithm or procedure to solve a problem, Based on the data collection that has been made, the author describes the analysis of the flowchart system framework as follows (Rahman & Santoso, 2015):

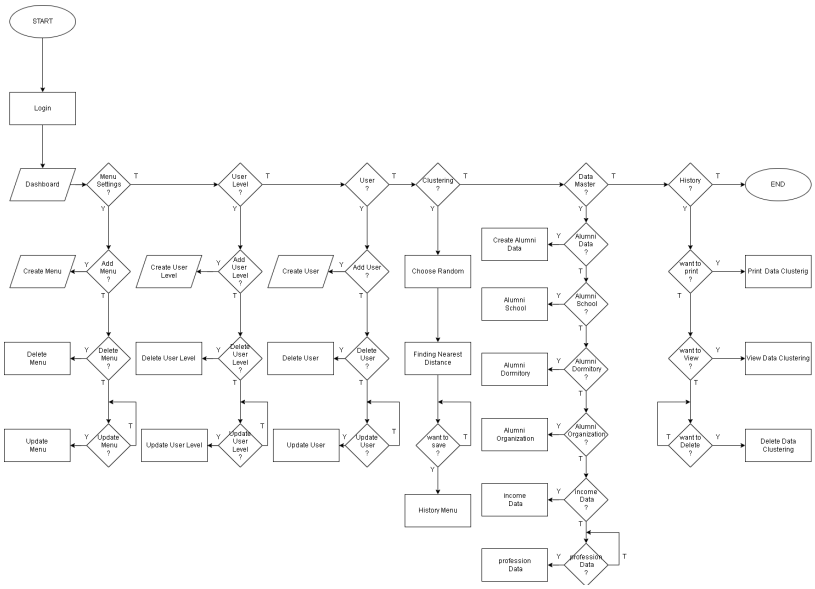

Figure 1. System Flowchart

DFD (Data Flow Diagram) Design

Data Flow Chart (DFD) is a modeling tool that describes the system as a network of functional processes that are connected with data flow either manually or computerized, the following is a data flow chart from the Alumni Web(Irrawan et al., 2019) :

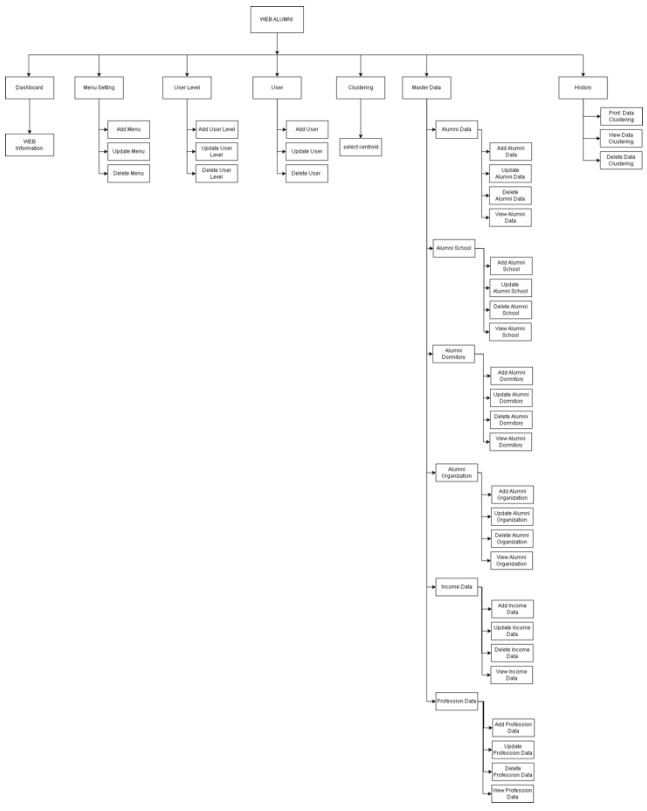

Figure 2. Data Flow Diagram.

RESULTS AND DISCUSSION

This research is the implementation of the system and the stage of application of the design that has been made. In this section, we will discuss the process of implementing the K-mean clustering algorithm on alumni web information system data.

Based on the data collection that has been made, the author provides examples of data The assumption of the work can be seen in Figure 3. And data on income assumptions can be seen in Figure 4. As for the alumni data that has been assumed to be in Figure 5. Random centroid selection for distance calculations is in Figure 6.

From the data analysis, the author gives the results of success with poverty indicators (Sitepu, 2012)based on 3 groups, C1 = Success, C2 = Quite Successful, C3 = Not Successful.

NO.	PEKERJAAN ALUMNI	ASUMSI
1	Belum/ Tidak Bekerja	1
2	Mengurus Rumah Tangga	2
3	Pelajar/ Mahasiswa	3
...	...	...
...	...	...
87	Biarawati	87
88	Wiraswasta	88

Figure 3. Job Data Assumption.

NO.	PENGHASILAN ALUMNI	ASUMSI
1.	< 500.000	1
2.	500.000-1.000.000	2
3.	1.000.000-3.000.000	3
4.	> 5.000.000	4

Figure 4. Income Data Assumption.

No.	Nama	Perkerjaan Alumni	Penghasilan
1	Nurul Annisa Sholehah	2	2
2	Isna Septiani Mariam	2	2
3	Zhaker	1	2
...	...	...	...
...	...	...	...
400	Putri Kusuma	1	4

Figure 5. Alumni Data.

Nama	Id Pekerjaan	Id Penghasilan	Centeroid
Ari	2	3	m1
Dian	1	3	m2
MUHAMMAD FADIL ZULFIKAR	2	1	m3

Figure 6.Centeroid Random.

The equation with the same calculation is applied to 400 data to obtain the distance of each data on each cluster, as in Figure 7 using the euclidian distance formula with the calculation

$$D(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \cdots (x_{ki} - x_{kj})^2}$$

Where:

D(i,j) = Distance of data to i to the center of the cluster j
x_ki = Data to i data attribute to k
x_kj = Center point to j on the attribute to k

Calculation of the distance of data to the centroid random

$d_{(m1)} = \sqrt{(2-2)^2 + (3-2)^2} = 1$
$d_{(m2)} = \sqrt{(1-2)^2 + (3-2)^2} = 1{,}414213562$
$d_{(m3)} = \sqrt{(2-2)^2 + (1-2)^2} = 1$

First interaction:

$$BCV = \sqrt{(2-1)^2 + (3-3)^2 + (2-2)^2 + (3-1)^2 + (1-2)^2 + (3-1)^2} = 5.236067977$$

WCV Total = 485
Rasio = 0.010796016

Where:

$d_{(m1)}, d_{(m2)}, d_{(m3)}$ = Distance of data to centroid
clusters that are followed= Is a variable from the distance of the data to the centroid the least value
closest distance = Variable minimum value from data distance to centroid
BCV = Is the calculation of the root rank of all centroids random
WCV = Power root of the nearest distance value
WCV Total = Is the total of the WCV
Ratio = Is the value of BCV divided by the total WCV value

NO.	NAMA	JARAK DATA KE CETEROID			CLUSTER YANG DI IKUTI	JARAK TERDEKAT	WCV
		m1	m2	m3			
1	Nurul Annisa Sholehah	1	1,4142 13562	1	1	1	1
2	Isna Septiani Mariam	1	1,4142 13562	1	1	1	1
3	Zhaker	1,414 2135 62	1	1,414213 562	2	1	1
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
400	Putri Kusuma	1,414 2135 62	1	3,162277 66	2	1	1
	Total	450,9 9764 91	474,14 59327	723,6524 115			485
	Rata-Rata	1,307 2395 63	1,3743 36037	2,097543 222			

Figure 7. First Cluster.

The following is the calculation of the second cluster with a new centroid the first is as follows:

Id Pekerjaan	Id Penghasilan	Centeroid
2.702703	2.864865	m1
1	3.037037	m2
1.8	1	m3

Figure 8. First New Centeroid

$$d_{(m1)} = \sqrt{(2{,}702703-2)^2+(2{,}864865-2)^2} = 0{,}914536449$$
$$d_{(m2)} = \sqrt{(1-2)^2+(3{,}037037-2)^2} = 2{,}252875011$$
$$d_{(m3)} = \sqrt{(1{,}8-2)^2+(1-2)^2} = 1{,}562049935$$

Second interaction :

$$BCV = \sqrt{\begin{matrix}(2.702703-1)^2+(2.864865-3.037037)^2+(2.702703-1.8)^2 \\ +(2.864865-1)^2+(1-1.8)^2+(3.037037-1)^2\end{matrix}} = 5.971739919$$

WCV Total = 374.9603225
Rasio = 0.015926325

After four calculations and the centroid has not changed, the cluster results are found, the following the calculation from the 4th cluster is as follows :

Id Pekerjaan	Id Penghasilan	Centeroid
2.324324	3.248649	m1
1.518519	2.511111	m2
1.8	1	m3

Figure 9.Third new centeroid

$$d_{(m1)} = \sqrt{(2{,}324324-2)^2+(3{,}248649-2)^2} = 1{,}419739788$$
$$d_{(m2)} = \sqrt{(1{,}518519-2)^2+(2{,}511111-2)^2} = 1{,}567170044$$
$$d_{(m3)} = \sqrt{(1{,}8-2)^2+(1-2)^2} = 1{,}562049935$$

Fourth Interation:

$$\sqrt{\begin{matrix}(2.324324-1.518519)^2+(3.248649-2.511111)^2+(2.324324-1.8)^2\\ +(3.248649-1)^2+(1.518519-1.8)^2+(2.511111-1)^2\end{matrix}} = 4.938448407$$

= 4.938448407
WCV Total = 429.4510719
Rasio = 0.011499444

No.	Nama	Jarak data ke ceteroid			cluster yang di ikuti	Jarak Terdekat	WCV
		m1	m2	m3			
1	Nurul Annisa Sholeha h	1,419739788	1,567170044	1,562049935	1	1,41973979	2,015661
2	Isna Septiani Mariam	1,419739788	1,567170044	1,562049935	1	1,41973979	2,015661
3	Zhaker	0,408671528	0,686175461	2,009975124	1	0,40867153	0,167012
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
400	Putri Kusuma	2,609646654	1,597597648	0,8	3	0,8	0,64
	Total	501,994392	449,365975	723,6029028			429,4511
	Rata-Rata	1,302510072	1,302510072	2,097399718			

Figure 9.Fourth Cluster

Result

At this stage, the overall system development process is carried out based on the results of the designs that have been made and implemented into the system Web. Pages built on In this system, the following is the interface page display:

- Log-in Page Display Results

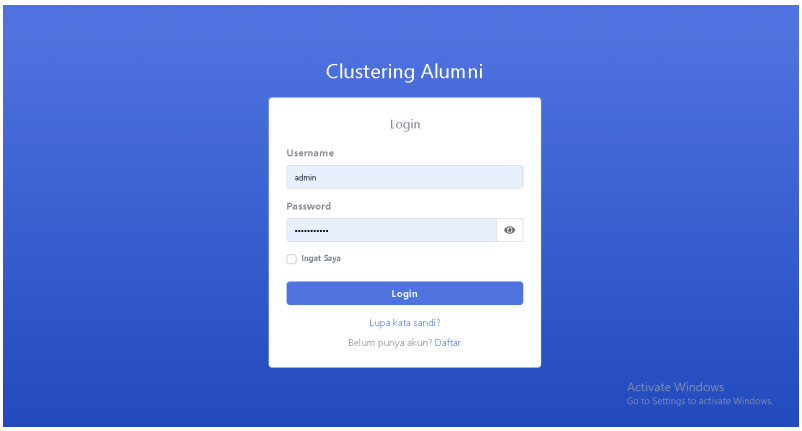

Figure 10. Log-in Page

The Log-in page is the first page that will be displayed when accessing this website, on this login page, enter based on the user level, and if you enter the wrong password, you can contact the admin. Following an interface design can be seen in (Figure 10).

- Main Menu Page Display Results (Dashboard)

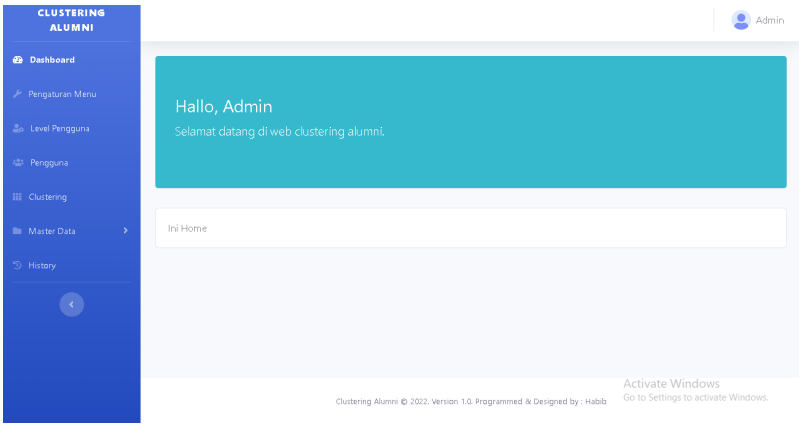

Figure 11. Main Menu Page

This page has 6 menus, namely menu settings, user level, User, alumni data, clustering, and master data while the master data menu has 5 sub-menus, namely school data, cottage data, community organization data, job data, and income data. The main page Interface Design can be seen in (Figure 11).

- Menu Settings Page

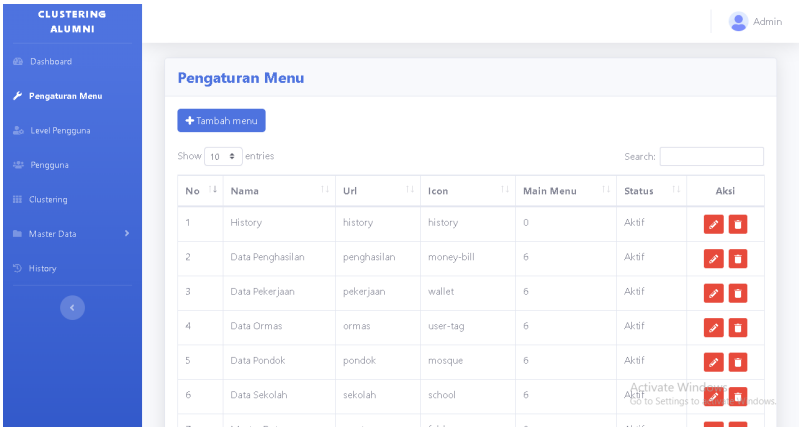

Figure 12. Menu Settings Page

On this page there are settings to add or view menus to be added, admins can add, edit, and delete menus that will be adjusted to the needs of the website, and on this page, there are relationships between tables of the entire menu. The Settings menu page can be seen (figure 12).

- User Level Page

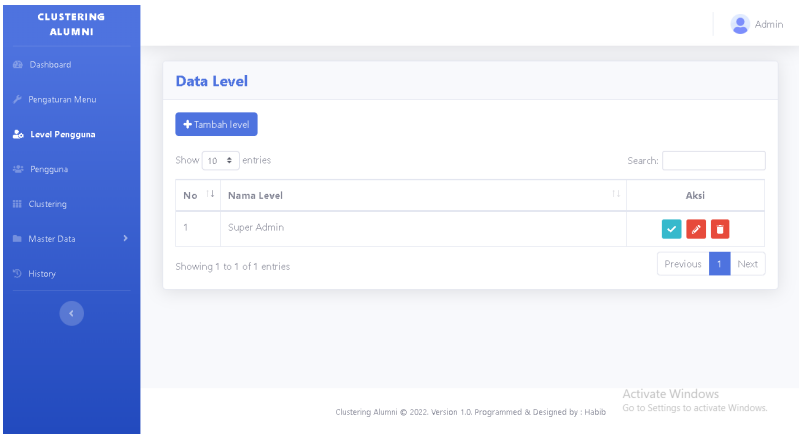

Figure 13. User Level Page

The following is a user-level page on the menu of this page admins can add, edit, and delete user levels, on this page is a page to determine user logins that distinguish between admins and users can be seen in (Figure 11).

- User Menu Page

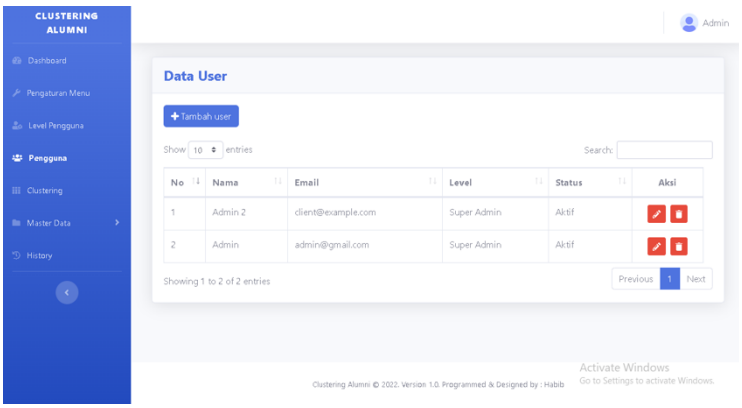

Figure 14. User Menu Page

This User menu page is a display that contains information about how many admins will access the dashboard and on this page, the Super admin can determine how many admins to enter to manage the alumni website as shown in (Figure 14).

- Clustering Menu Page

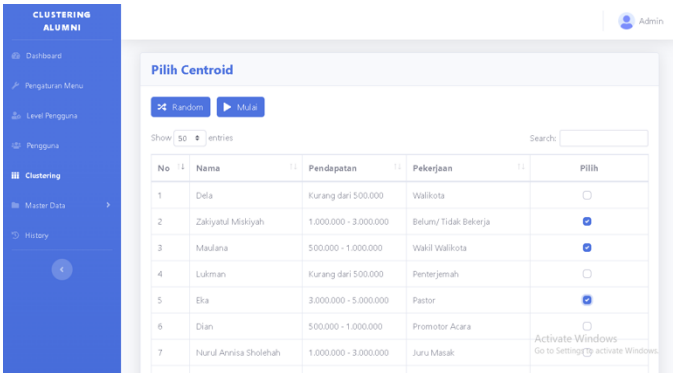

Figure 15. Clustering Menu Page

This clustering menu page is a page that contains data about alumni who will be randomly clustered, on this page the user must choose three alumni data at random, or directly click on the random menu and they will randomly choose the data to be in the cluster. as shown in (Figure 15).

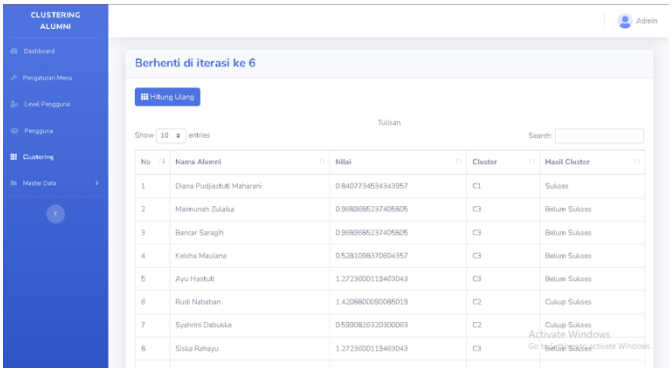

Figure 16. Results From Clustering Page

This clustering results menu page is a page that contains the results of alumni data after the centroid calculation is carried out, on this page the calculation is carried out until the centroid is the same as determining the distance from each calculated literacy. Inside (Figure 16).

- History Menu Page

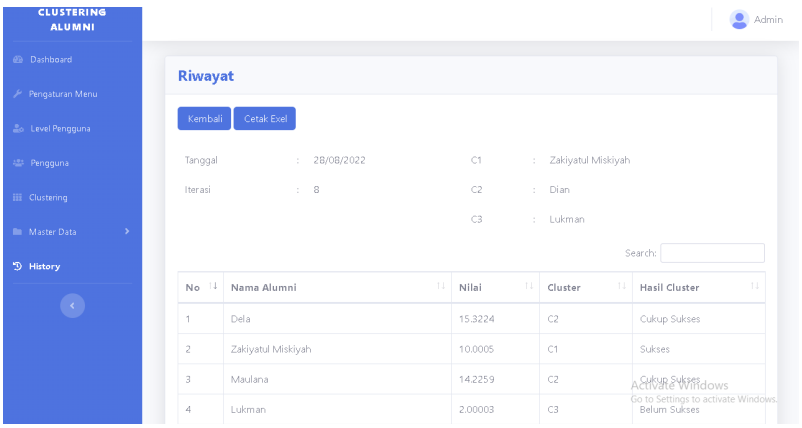

Figure 16. History Menu Page

This history menu page is a page that contains historical data from the clustering menu, on this page you can also download data that has been clustered and can delete data that is no longer needed. as shown in (Figure 16).

CONCLUSION

Based on the research that has been carried out, a conclusion can be drawn that the implementation of this algorithm is used to make it easier for alumni and related parties to find out information easily, and see detailed alumni data anytime and anywhere. As well as making it easier for caregivers to cluster without having to wait a long time in finding data on alumni of the Bahrul Ulum Islamic Boarding School Foundation.

REFERENCES

Ali, A. (2019). Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *19*(1), 186–195. https://doi.org/10.30812/matrik.v19i1.529

Destari, D. (2015). Peningkatan Kualitas Program Studi Pendidikan Bahasa Inggris Jurusan Tarbiyah Sekolah Tinggi Agama Islam Negeri (STAIN) Samarinda Berbasis Akreditasi. *Fenomena*, *7*(1), 45. https://doi.org/10.21093/fj.v7i1.265

Dhuhita, W. (2015). Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita. *Jurnal Informatika Darmajaya*, *15*(2), 160–174.

Irrawan, S. N., Simanjuntak, R. A., & Yusuf, M. (2019). ISSN : 2338-7750 Institut Sains & Teknologi AKPRIND Yogyakarta Jurnal REKAVASI ISSN : *Jurnal REKAVASI*, *7*(1).

Nugroho, K. A., Munawaroh, M., & Hariono, T. (2020). Rancang Bangun Dan Implementasi Web Service Pada Sistem Informasi Akademik Berbasis Mobile Studi Kasus Di Unwaha. *SAINTEKBU*, *12*(1), 38-47.

Putrawansyah, F., & Dewi, N. C. (2021). *Implementasi Algoritma K-Means dengan menggunakan Metode Profile Matching pada Alumni STT Pagar Alam Implementation of the K-Means Algorithm using the Profile Matching Method for Alumni of STT Pagar Alam*. *0*(02), 47–53.

Rahman, F., & Santoso. (2015). Aplikasi pemesanan undangan online. *Sains Dan Informatika*, *1*(2), 78–87.

Sitepu, A. (2012). *Karakteristik Keluarga Menurut Peringkat Kemiskinan: Studi Pendahuluan untuk Perumusan Kriteria Fakir Miskin*. Informasi. https://ejournal.kemsos.go.id/index.php/Sosioinforma/article/viewFile/930/490