Jurnal Teknik Informatika (JUTIF)
Vol. 3, No. 5, Oktober 2022, hlm. 1189-1194

DOI: https://doi.org/10.20884/1.jutif.2022.3.5.296
p-ISSN: 2723-3863
e-ISSN: 2723-3871

DATA CLUSTERING ON TEACHING AND LEARNING EVALUATION APPLICATION OVER PANDEMIC ERA

Arie Vatresia[*1], Yosan Frediantо[2], Ashahar Johar T[3]

[1,2,3]Program Studi Informatika, Fakultas Teknik, Universitas Bengkulu, Indonesia
Email: [1]arie.vatresia@unib.ac.id, [2]sevenick.17@gmail.com, [3]asahar.johar@unib.ac.id

(Naskah masuk: 25 April 2022, Revisi : 27 April 2022, diterbitkan: 24 Oktober 2022)

Abstract

SIEPEL is an evaluation application for teaching and learning process in University of Bengkulu. It is mandatory for every student to fill the questionnaire before they can see the marking value for each subject each semester. This survey was designed to meet the requirements and expectations of students as educational service for customers. This data is very important to improve the quality of teaching and learning process for further policy and decision maker. However, the analysis of the data remains an open question as the size and the distribution of the data is become some issues to process the analysis. Here, we showed the new approach to analysis the data using K-Means Clustering to see the better distribution and understanding over the evaluation data. This paper used elbow method to find the best number of clusters to be implemented on the algorithm approach which results in four clusters of satisfaction values (unsatisfied, less satisfied, satisfied, and very satisfied). The result of this analysis was published based on website system to show the visualization of analysis. Furthermore, this research showed that the average value of evaluation result for 4 semester was very satisfied 6.50%; satisfied 43.89%; less satisfied 44.26%; and not satisfied 5.36%. The value of vary satisfied students was dropped from 20.47% to 0.12% by 2 years and the value for less satisfied was increased from 27.64% to 66.32%. This term was happened because of the pandemic era and the change on the process of learning and teaching on University of Bengkulu.

Keywords: *Data Visualization, K-Means Clustering, Lecturer Performance, Student Satisfaction.*

1. INTRODUCTION

SIEPEL UNIB is a Bengkulu University learning evaluation information system that was established to fill in the gaps in the evaluation of the current semester's courses. Students can use it to evaluate the lecturer in charge of the course's performance in the classroom during teaching and learning activities. It was created in 2018 and continues to be used today. The measurement of student satisfaction, which is one part of the assessment of educational service quality, has significant implications for the educational system's long-term sustainability [1]. Data visualization is concerned with the design, development, and implementation of computers [2]. Previously, research on the visualization of student satisfaction data against lecturers was conducted by Candra [3] who examined the Prediction of Student Satisfaction on Service Levels using the C4.5 Decision Tree Algorithm, Research on Application of Kmeans Algorithm for Clustering Lecturer Assessment Based on Student Satisfaction Index by khusniati [4] Research on Student Satisfaction on Lecturer Performance by Ruslan [5], Research on Analysis of Student Satisfaction on Lecturer Performance by Sulatri [6], Research on Student Satisfaction Levels on Performance of Lecturer Teaching and Learning Process by Paly [7], Research on Visualization of Patient Laboratory Examination Results by Ardy [8], Research on Interactive Data Visualization of Open Data by Syaripul [9], Research on Sales and Production Data Visualization by Aryanti [10], Research on Data Visualization in Library Management Systems by Saputra [11], and Research on Design and Development of Data Visualization on Internal Research Funds and Grants by Loka [12].

The basic goal of data visualization is to show data in such a way that it may clearly and descriptively express information [13]. Users can utilize effective visualization to help them analyze and reason about facts and evidence [14]. Data visualization, according to Jer Thorp, is "anything that shows a narrative or picture so that individuals can grasp things more deeply" [15]. The elbow approach is used to discover the ideal cluster by first picking the cluster value and then adding the cluster value to be utilized as a data model [16]. For example, from K = 3 to K = 4, there is a drastic decrease in the form of an angle at K = 4. K = 4 is the optimal cluster value of k [17]. The usage of the kmeans method for clustering in this essay has numerous explanations, which are as follows: The first reason is the ease with which the kmeans algorithm can be implemented and executed [18, 19]. The second reason is that clustering is a data segmentation technique that can help foresee and analyze certain business problems [20]. The third reason is that K-means is the most

basic and widely used clustering algorithm [21]. This project will collect data from SIEPEL UNIB and conduct research on visualization of student satisfaction data on the performance of lecturers at Bengkulu University.

2. METHODS

The research was conducted based on the research stages below, this is also adjusted to the selected system development method [15]. The first stage is the initial determination of a research, namely the research background. The second stage is to analyze the problem and the need for system design. At this stage, the analysis of the problem to be solved in the research is carried out so that it can find what is needed in the system [2]. Furthermore, the system design stage is carried out to describe the design of use case diagrams, activity diagrams, class diagrams, sequence diagrams, database design and the design of the system interface to be made. The implementation of the design results is to implement the design results into the PHP programming language [22] and the codeigniter framework along with the mysql database. The stages of system testing carried out are testing using the black box testing method. The last stage is drawing conclusions and suggestions on the system. This stage is carried out to find out the results that have been obtained during the research.

Since this system was built using SIEPELUNIB's existing database, the tables used in the database are briefly shown. This project used 4 different tables in the SIEPEL UNIB database: tb_semester, tb_rekap_dosen, tb_rekap_prodi, and tb_rekap_fakultas. Please consult the table below for more information.

Table 1. tb_semester

Name	Type	Size	Description
id_semester	Int	11	Primary Key
nm_semester	Varchar	50	
status	Int	1	

Table 1 is the tb_semester table offers information about the semesters that have been filled out on the SIEPEL website's evaluation.

Table 2. tb_rekap_dosen

Name	Type	Size	Description
id	Int	11	Primary Key
id_semester	Int	11	Foreign Key
nip	Varchar	18	
nama	Varchar	50	
id_prodi	Varchar	10	
nm_prodi	Varchar	50	
id_fakultas	Varchar	2	
nilai_pedagogik	Double		
nilai_profesional	Double		
nilai_kepribadian	Double		
nilai_sosial	Double		
nilai	Double		
keterangan	Varchar	50	
jumlah_data	Int	11	
waktu_post	Datetime		
waktu_edit	Datetime		

Table 2 is the tb_rekap_dosen table offers information about each lecturer's score on four assessment indicators: pedagogic values, professional values, personality values, and social values. The four values are derived from the assessments that each student completes each semester.

Table 3. tb_rekap_prodi

Name	Type	Size	Description
id	Int	11	Primary Key
id_semester	Int	11	Foreign Key
id_prodi	Int	11	
nm_prodi	Text		
jenjang	Varchar	32	
id_fakultas	Int	11	
nilai	Float		
jumlah_kelas	Varchar	32	
jumlah_dosen	Varchar	32	
jumlah_mhs	Varchar	32	

Table 3 is the tb_rekap_prodi table, which offers information regarding Bengkulu University's study programs. Because the values utilized pertain to the values in the tb rekap dosen table and are used to determine the clusters obtained by each study program, the value column is not used here.

Table 4. tb_rekap_fakultas

Name	Type	Size	Description
id	Int	11	Primary Key
id_semester	Int	11	Foreign Key
id_fakultas	Int	11	
nm_fakultas	Text		
nm_singkat	Text		
nilai	Float		
jumlah_kelas	Varchar	32	
jumlah_dosen	Varchar	32	
jumlah_mhs	Varchar	32	

Table 4 is the tb_rekap_fakultas table, which contains information on Bengkulu University's faculties. Because the values utilized pertain to the values in the tb rekap dosen table and are used to determine the clusters obtained by each faculty, the value column is not used here.

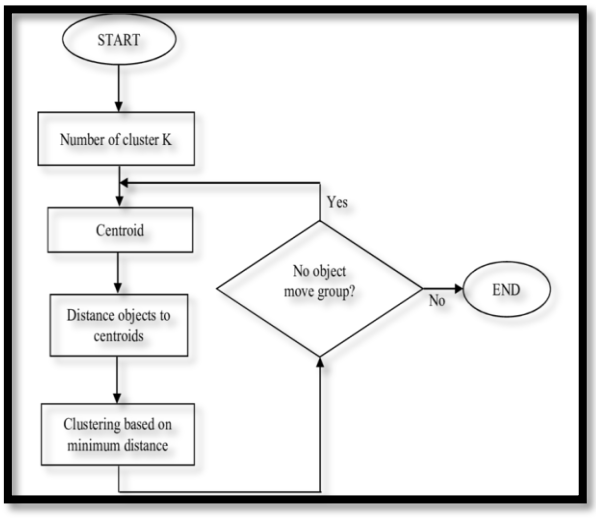

Figure 1 K-means Flowchart

Description of the system flow chart process and its coding based on the system flow chart in Figure 1 The first step is to figure out how many k-clusters there are. The elbow approach can be used to calculate the ideal number of k-clusters. The elbow method is a way of generating data in order to figure out how many clusters there are. The most effective method is to examine the percentage of the results of a comparison of the number of clusters that will create an angle at a place.

3. DISCUSSION

```
kmeans_kwargs = {"init": "random", "n_init": 10,
   "max_iter": 300, "random_state": 1234}
sse = []
for k in range(1, 11):
kmeans = Kmeans(n_clusters=k, **kmeans_kwargs)
kmeans.fit(df.loc[:, ['nilai_pedagogik',
                      'nilai_profesional',
                      'nilai_kepribadian',
                      'nilai_sosial']])
sse.append(kmeans.inertia_)
plt.style.use("fivethirtyeight")
plt.plot(range(1, 11), sse)
plt.xticks(range(1, 11))
plt.xlabel("Number of Clusters")
plt.ylabel("SSE")
plt.savefig('elbow-method.png')
plt.show()
kl = KneeLocator(range(1, 11), sse, curve="convex",
direction="decreasing")
print("jumlah cluster dri elbow method: ")
print(kl.elbow)
```

Figure 2 Elbow Calculation

The Elbow method is a method to determine the right number of clusters through the percentage of the comparison between the number of clusters that will form an angle at a point. If the value of the first cluster with the value of the second cluster gives the angle in the graph or the value has decreased the most, then the number of cluster values is the right one. To get a comparison is to calculate the Sum of Square Error (SSE) of each cluster value. Because the larger the number of cluster K values, the smaller the SSE value will be. This method provides ideas by selecting the cluster value and then adding the cluster value to be used as a data model in determining the best cluster. Python code was used to discover the best k value using the elbow approach, as illustrated in Figure 2. Figure 2 initializes two variables, the kmeanskwargs and sse variables. These remain empty. Then, repeating the value of k in the range of values k1 to k11, we can experiment with the elbow approach from k = 1 to k = 11. The kmeans variable is then defined. Following that, we establish the data's location and define the value of SSE.The elbow graph can then be displayed using a plot, and the best elbow value can be chosen, as well as the number of clusters or the optimal k value, by utilizing the kneelocator function at a distance of k1 to k11. When the preceding code is run, the elbow method returns a graphical depiction along with the number of clusters. The results demonstrate that four is the best number of k, as seen in Figure 3.

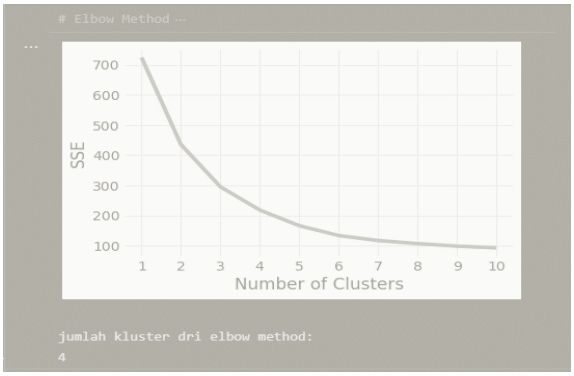

Figure 3 Elbow Graph Results

Figure 3 shows that various K values steadily fall until the results of the K values are steady. Figure 4 shows that from K = 3 to K = 4, there is a significant drop in the appearance of an elbow at K = 4, indicating that K = 4 is the best cluster value for k.In the second step, the second process can determine the centroid of point k (the center of the cluster). Here's how to use code like Figure 4 to run the process of getting the centroid or midpoint in Python.

```
dfCp1 = df.copy(deep=True)
km = Kmeans(init="random", n_clusters=4,
max_iter=300, random_state=1234)

km.fit(dfCp1.loc[:, ['nilai_pedagogik',
'nilai_profesional',
'nilai_kepribadian',
 'nilai_sosial']])

print("Centroid cluster : ")
print(km.cluster_centers_)
```

Figure 4 Coding Determines the Centroid Value

After determining that k4 is the optimal k value, configure n clusters equal to the four declarative variables from the data frame to initialize the km variable and set the positions of the four dates to calculate at the Figure 4. Finally, the centroid value is printed or displayed. The code will give you the result of centroid result of the clusters value. Then the centroid values are printed or displayed.

```
[[4.09649821  4.10295193 4.09318161 4.0763489]
 [4.33682416  4.3257771  4.33457398  4.28287591]
 [4.74216667  4.75659728 4.76495102 4.68515442]
 [3.56273571  3.62426161 3.53510268 3.40427411]]
```

Figure 5. Centroid Results

Figure 5 shows the results, with each of us receiving four centroid values for four different sorts

of values: pedagogic values, professional values, personality values, and social values. The third step is to compute the distance between each object and each cluster's centroid. The inertia value can be printed on python. The km.inertia function is used in Figure 6 to compute the distance between the item and the centroid. The results of the inertia value, namely the total value of the distance between each point and the nearest centroid. The fourth step is to group or allocate each object to the centroid that is closest to it.

A new label called "Cluster" in the dfcp1 data frame. This helps to group the clusters received from each lecturer. And, as seen in Figure 9, the dfcp1 data frame is displayed after that. The outcome of grouping the clusters of each id and assessment is shown in Table 5.

Table 5. Clusters

Id	Pedagogik	Profesional	Kepribadian	Sosial	Cluster
1	4.240	4.162	4.246	4.083	0
2	3.297	3.586	3.231	3.104	3
3	4.453	4.331	4.407	4.321	1
4	4.506	4.367	4.445	4.382	1
5	3.043	3.112	3.311	3.247	3
6	4.651	4.512	4.689	4.601	2

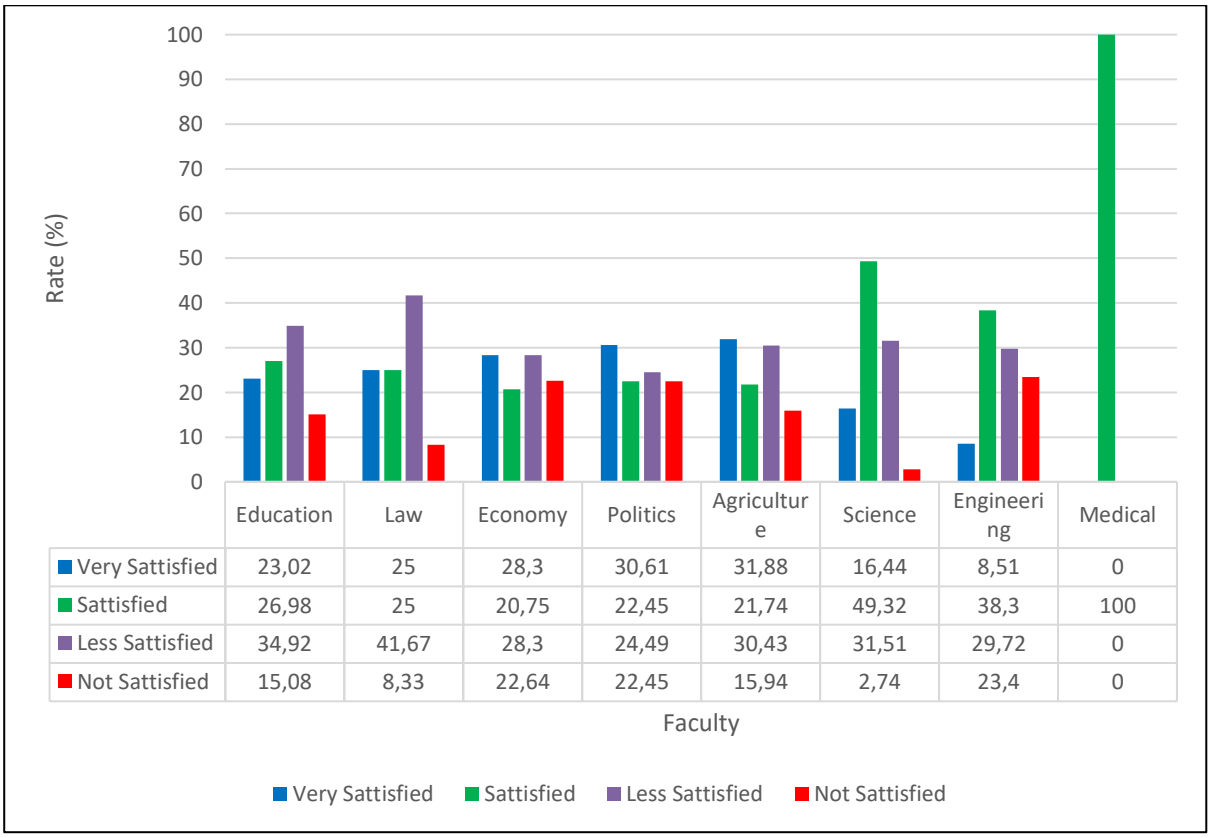

Figure 6. Clusters Distibution For Each Faculty

Figure 6 shows the distribution of sattisfactory rate for every faculty in the second term of the year 2018/2019. Although the rate of not sattisfied (colour red) was 12%, we can see that the very sattisfactory rate in this time was 22% and sattisfied rate was 37% that were dominated the rate of education and learning rate. There are variation on distribution of sattisfaction rate, including for personal option. There is one phenomenom that is interesting to be seen. One lectures can gets dissatisfied grades on a regular basis and the other consistently gets satisfactory grades. This paper found that one lecture can received low scores on each parameter, including pedagogic values, professional values, personality values, and social values. In another hand, there was another lecture who received a satisfied category value every semester, according to the findings of the k-means calculation who did excellent performance. The four assessments were rather comprehensive, including pedagogic values, professional values, personality values, and social values. This study also explored the rate for faculty as one of the analytic sides that can be used for further policy. Figure 7 showed the distribution of satisfaction over Engineering Faculty for 5 semesters. We can see that in the second semester of 2018/2019, we still find 40% of students were satisfied with the learning teaching activity in Faculty of Engineering. The value of satisfied rate was increased to 53% in the beginning of year 2019, but then drop significantly in the end of year 2020 when the pandemic hit the university activity. Another phenomenom was found that the rate of less sattisfied rate were found significantly increase over the 5 semester from the end of 2018 to the end of 2020.

DOI: https://doi.org/10.20884/1.jutif.2022.3.5.296
p-ISSN: 2723-3863
e-ISSN: 2723-3871

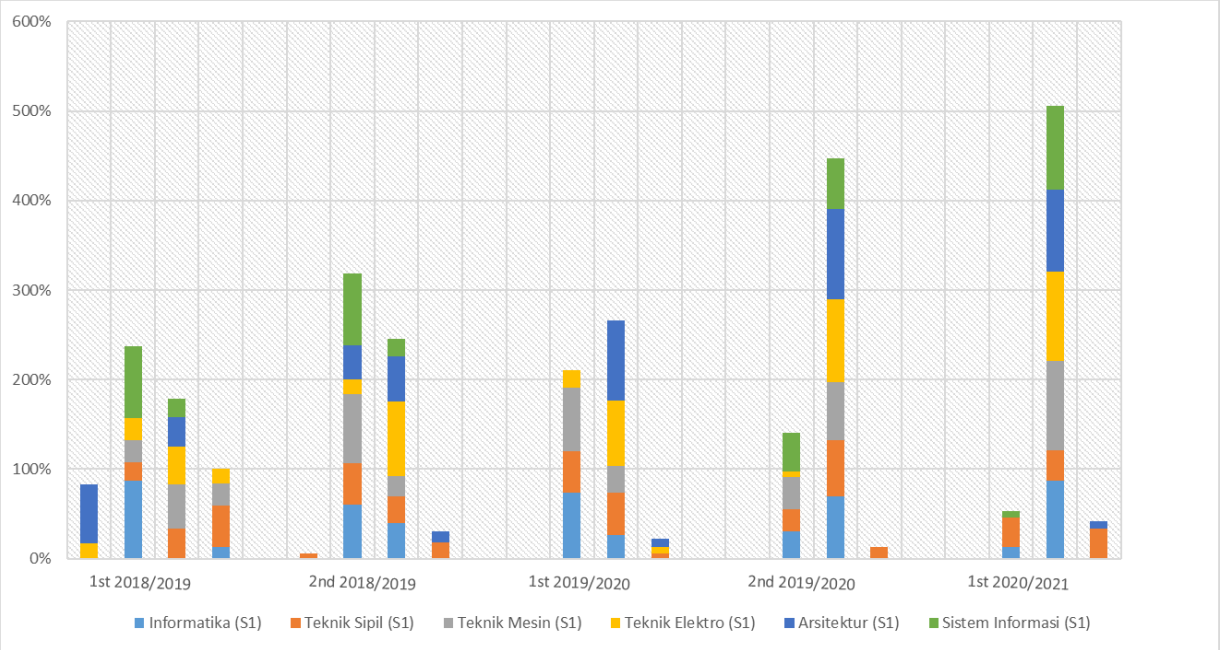

Figure 7 Engineering Evaluation rate distribution

This research utilised four assessment indicators, namely pedagogic assessment, professional assessment, personality assessment, and social assessment, and after calculating the results using Kmeans clustering on the four assessments, the student satisfaction with lecturers at Bengkulu University is as follows: The total score summary results for the odd semester of the 2018/2019 academic year showed that 20.47 percent of students were very satisfied, 38.07 percent of students were satisfied, 27.64 percent of students were less satisfied, and 13.82 percent of students were unsatisfied. In the even semester of the 2018/2019 academic year, 3.40 percent of students were very satisfied, 51.04 percent of students were satisfied, 41.89 percent of students were less satisfied, and 3.67 percent of students were unsatisfied, according to the overall score review. In the odd semester of the 2019/2020 academic year, 2 percent of students were very satisfied, 55.44 percent of students were satisfied, 41.17 percent of students were less satisfied, and 1.39 percent of students were less satisfied, according to the overall score summary. In the even semester of the 2019/2020 academic year, 0.12 percent of students were very satisfied, 31.02 percent of students were satisfied, 66.32 percent of students were less satisfied, and 2.54 percent of students were unsatisfied, according to the overall score review.

4. CONCLUSION

This research showed the significant impact of pandemic affected the process of teaching and learning activity over 5 semesters over University of Bengkulu. The clusters was succeed to devide the satisfaction rate into 5 clusters acording to performance of lecturers over every semesters. However, this result migh still need improvement as the performance of K-Means only processed numeric value of sattisfactory rate while still neglecting the naration value. We encourage the bigger data size for further research with advanced data structured to reduce time complexity over the processing time. Furthermore, this study was only used 4 tables from Siepel database, so it is hoped that further research can also use other tables for another significant result. Visualization of data in this study only uses two-dimensional graphs further research can also use three-dimensional graphics for better understanding of the data distribution.

REFERENCES

[1] L. Herawati dan D. Setialesmana, “Kepuasan Mahasiswa berdasarkan Kinerja Dosen dalam Proses Perkuliahan di FKIP Jurusan Matematika Universitas Siliwangi,” Jurnal Siliwangi, vol. 2, no. 2, pp. 131-136, 2016.

[2] M. N. O. Sadiku, A. E. Shadare, S. M. Musa dan C. M. Akujuobi, “Data Visualization,” International Journal of Engineering Research And Advanced Technology(IJERAT), vol. 2, no. 12, pp. 11-16, 2016.

[3] B. A. CandraP, “Prediksi Kepuasan Mahasiswa Terhadap Tingkat Pelayanan menggunakan Algoritma C4.5(Decision Tree),” Infotek : Jurnal Informatika dan Teknologi, vol. 1, no. 1, pp. 32 - 39, 2018.

[4] M. Khusniati, M. A. Muslim dan A. Nurzahputra, “Penerapan Algoritma K-Means Untuk Clustering Penilaian Dosen

Berdasarkan Indeks Kepuasan Mahasiswa," Techno.COM, vol. 16, no. 1, pp. 17-24, 2017.

[5] Ruslan, "Kepuasan Mahasiswa Terhadap Kinerja Dosen," Jurnal Ilmu Pendidikan, vol. 17, no. 3, pp. 230-237, 2010.

[6] T. Sulastri, "Analisis Kepuasan Mahasiswa Terhadap Kinerja Dosen," Jurnal Ilmiah Ekonomi Manajemen dan Kewirausahaan, vol. 10, no. 2, pp. 167-184, 2016.

[7] B. Paly, "Tingkat Kepuasan Mahasiswa Terhadap Kinerja Proses Belajar Mengajar (PBM) Dosen," Jurnal Al Hikmah, vol. 15, no. 2, pp. 250-262, 2014.

[8] Y. Ardy dan H. Budi, "Visualisasi Hasil Pemeriksaan Laboratorium Pasien Studi Kasus: Parahita Diagnostic Center," JUISI, vol. 1, no. 2, pp. 142-150, 2015.

[9] N. A. Syaripul dan A. M. Bachtiar, "Visualisasi Data Interaktif Data Terbuka Pemerintah Provinsi DKI Jakarta," Jurnal Sistem Informasi (Journal of Information System), vol. 12, no. 2, pp. 82-89, 2016.

[10] D. Aryanti dan J. Setiawan, "Visualisasi Data Penjualan dan Produksi PT Nitto Alam Indonesia Periode 2014-2018," ULTIMA InfoSys, vol. 9, no. 2, pp. 86-91, 2018.

[11] D. F. Saputra, "Visualisasi Data Di Sistem Manajemen Perpustakaan," Jurnal Perpustakaan Pertanian, vol. 26, no. 2, pp. 82-86, 2017.

[12] W. I. Loka dan F. Natalia, "Perancangan dan Pembuatan Visualisasi Data Dana Penelitian Internal dan Hibah Dikti LPPM Universitas Multimedia Nusantara," ULTIMA InfoSys, vol. 10, no. 1, pp. 61-68, 2019.

[13] F. Vitaly, Data Visualization and Graphics, USA: Monday Inspiration, 2008.

[14] B. Fry, Visualizing data, USA: O'Reilly Media, 2008.

[15] J. Thorp, Beautiful Visualization: Looking at Data through the Eyes of Experts, California: O'Reilly Media, 2010.

[16] T. S. Madhulatha, "An Overview On Clustering Methods," IOSR Journal of Engineering, vol. 2, no. 4, pp. 719-725, 2012.

[17] T. M. Kodinariya dan P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," International Journal of Advance Research in Computer Science and Management Studies, vol. 1, no. 6, pp. 90-95, 2013.

[18] Syafnidawaty, "K-Means Clustering," Universitas Raharja, 19 April 2020. [Online]. Available: https://raharja.ac.id/2020/04/19/k-means-clustering/. [Diakses 20 9 2021].

[19] D. S. Saputri, G. M. Putra dan M. F. Larasati, "IMPLEMENTATION OF THE K-MEANS CLUSTERING ALGORITHM FOR THE COVID-19 VACCINATED VILLAGE IN THE UJUNG PADANG SUB-DISTRICT IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK DESA TERVAKSINASI COVID-19 PADA KECAMATAN UJUNG PADANG," JUTIF, vol. 3, no. 2, pp. 261-267, 2022.

[20] E. Irwansyah, "Clustering," Universitas Bina Nusantara, 9 Maret 2017. [Online]. Available: https://socs.binus.ac.id/2017/03/09/clustering/. [Diakses 1 12 2020].

[21] T. Alfina, B. Santosa dan A. R. Barakbah, "Analisa Perbandingan Metode Hierarchical Clustering, K-Means Dan Gabungan Keduanya Dalam Membentuk Cluster Data," Jurnal Teknik Pomits, vol. 1, no. 1, pp. 1-5, 2012..