UNS
UNIVERSITAS
SEBELAS MARET

PROCEEDING

ICoSMEE

THE 2ND INTERNATIONAL CONFERENCE ON SCIENCE, MATHEMATICS, ENVIRONMENT, AND EDUCATION

Innovative Research in Science, Mathematics, Environment and Education for Sustainable Development

26-28 July 2019

Science and Math Education Department
Faculty of Teacher Training and Education
Universitas Sebelas Maret
Jl. Ir. Sutami 36A Kentingan
Surakarta - Jawa Tengah 57126

icosmee@mail.uns.ac.id
https://icosmee.uns.ac.id

PROCEEDING

ISBN:

The 2nd International Conference on Science, Mathematics, Environment, and Education (ICoSMEE)

"Innovative Research in Science, Mathematics, Environment and Education for Sustainable Life"

Surakarta, July 26 - 28 2019

Organizers
Science and Math Education Department
Faculty of Teacher Training and Education
Universitas Sebelas Maret
Jl. Ir. Sutami 36 A Kentingan Surakarta - Jawa Tengah 57126
website: http://icosmee-uns.org , email: icosmee@mail.uns.ac.id

PROCEEDING

The 2nd International Conference on Science, Mathematics, Environment, and Education (ICoSMEE)

"Innovative Research in Science, Mathematics, Environment and Education for Sustainable Life"

Editor : Dr. paed. Nurma Yunita Indriyanti, M.Si, M.Sc
Dr. Farida Nurhasanah
Murni Ramli, S.P., M.Si, Ed.D
Dimas Gilang R, M.Pd.

Reviewer : Dr. Sri Widoretno, M.Pd.
Dr. Muzzazinah, M.Si.
Murni Ramli, S.P., M.Si, Ed.D
Puguh Karyanto, S,Si, M.Si, Ph.D.
Dr.rer.nat. Sri Mulyani, M.Si.
Dr. paed. Nurma Yunita Indriyanti, M.Si, M.Sc
Lina Mahardiani, S.T., M.Sc., Ph.D
Dr. Elfi Susanti VH, S.Si.,M.Si.
Dr. Sri Retno Dwi Ariani, M.Si.
Dr. Endang Susilowati, M.Si.
Dr. Sarwanto, S. Pd, M.Si
Daru Wahyuningsih, S.Si, M.Pd
Dr. Mardiyana, M.Si.
Dr. Triyanto, S.Si, M.Si
Rubono Setiawan, M.Si.
Dr. Laila Fitriana, S.Pd, M.Pd.
Dr. Budi Usodo, M.Pd.
Dr. Sri Yamtinah, M.Pd
Dr. Meti Indrowati, S.Si., M.Si.

Desain Cover & Setting Lay Out:
Finly Khoirunnisa' A
Siti Nur Afifah
Rizka Fauzia Hanif

ISBN: 978-602-73159-6-9

Published by:
Chemistry Education Department
Faculty of Teacher Training and Education
Universitas Sebelas Maret
Jl. Ir. Sutami 36A Kentingan
Surakarta - Jawa Tengah 57126

LIST OF CONTENTS

LIST OF PRESENTER ON PARALLEL SESSION

MATHEMATICS AND MATHEMATICS EDUCATION

Clustering and Diversifying Student Spreads based on Environmental Effects assisted by WebGIS visualization

Achmad Fauzan, Aji Bani Ismaun, Muhammad Hasan Sidiq Kurniawan, Jaka Nugraha

Department of Statistics, Faculty of Mathemtics and Natural Science, Universitas Islam Indonesia

Correspondence: achmadfauzan@uii.ac.id

Abstract. The distribution of the number of students accepted in the study program environment is one of the studies in determining the promotion process of a study program. Considering the importance of the quality and original quantity of the prospective students, this study aims to examine clustering from prospective student regions based on the environmental influences of prospective students and visualize with the Website Geographic Information System (WebGIS). The object of this study from students Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia (FMIPA UII). FMIPA UII consists of five study programs, namely Statistics, Chemistry, Pharmacy, Chemical Analyst, and Chemical Education. The research data used are primary data obtained from random sampling. The research method in this research is Hierarchical Clustering. The variables in this study consisted of five components, namely the influence of the family (X1), siblings or neighbors (X2), successful people around prospective students (X3), teachers in the school (X4), and seniors during high school (X5). From the data analysis, each study program has several different clusters based on the origin of students from each study program. From the results of the clustering obtained, the number of students who choose study programs at FMIPA UII is not entirely proportional to the influence of the surrounding environment. As an example, in the undergraduate program in statistics, environmental factors influence the number of students, while in an undergraduate program in Chemistry it does not apply (in other words other factors are more influential prospective students choose Chemistry study programs). Cophenetic Correlation Coefficient is used as an evaluation of the cluster obtained. After clustering is done to facilitate visualization between clustering results and real data used by WebGIS. WebGIS visualization can be used continuously and facilitates the study program in evaluating future student admissions. Besides, WebGIS also helps study programs to a promotion, because it can be combined with the presence of profile videos, impressions of student messages, and other information to improve the quality and quantity of students.

Keywords: Cophenetic Correlation Coefficient, Hierarchical Clustering, WebGIS.

INTRODUCTION

The characteristics of prospective students in a college are influenced by various factors, one of the factors is the origin of the prospective student area. The more varied students are in the university environment, it can be used as an indicator that the dissemination of information or promotions carried out from study programs at a university is dynamic and optimal. Despite the vastness of the promotion that has been carried out, evaluation is needed for a more effective and efficient promotion, not only boils down to the quantity of promotion but the quality of promotion so that it is right on target. This is following the purpose of the evaluation, which is to influence the decisions needed for future programs, the need for modification and the provision of costs or benefits from the program that has been held [1]

This study focuses on grouping related to how much the environmental influence of prospective students in choosing their chosen study program at this time. The technique used is in the form of clustering techniques. Then, after the results of the clustering process are obtained, the data visualization process is continued. Good data visualization will be able to produce several benefits, including better decision making, increased collaboration/information dissemination, better ad-hoc data analysis, time savings, and increased profit/investment [2] Assisted by Website Geographic Information System (WebGIS), the visualization of the data generated can be more informative, dynamic, and continuous in the future.

The purpose of this study is to look at the pattern of grouping motivation of prospective students focused on the surrounding environment and then can be visualized in the form of online media in the form of WebGIS. From the pattern obtained later, it can be used as an evaluation of the study program in the promotion strategy and the dissemination of information related to the study program. Previous significant studies related to clustering included earthquake center clusters in Bengkulu Province using K-Means Cluster [3], hierarchical clustering and visualization using R for clinical research [4]. Meanwhile, Rana [5], uses the Hierarchical Clustering algorithm to evaluate the ability of students, and [6] related to clustering by using the K-Mean Algorithm to classify the data of prospective new students at Universitas Muhammadiyah Yogyakarta. Meanwhile, previous studies related to WebGIS include WebGIS application research in interactive development for urban management in Batna [7], application utilization from WebGIS to visualize coastal flood vulnerability and planning for its resilience in New Jersey [8]

The further impact is the optimization of the quality and quantity of students in the program. From superior seeds and optimal processes will produce quality alumni in each study program and strengthen the selling points of each study program.

RESEARCH METHOD

Data and Population

The research was conducted in three months (October to December 2018) in the Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia (FMIPA UII). The population in this study were students at the Faculty of Mathematics and Natural Sciences in all study programs (Statistics, Chemistry, Pharmacy, Chemical Analysts, and Chemistry Education) from the 2016/2017 class until 2018/2019. The data obtained is primary data from questionnaires to students. The research method used is hierarchical clustering which aims to classify the distribution of students who will be adjusted to the number of students. Then continued by WebGIS as a visualization of the results of clustering and descriptive statistics from each region. Samples were taken from the population using a random sampling method. The sample size used is presented in Eq 1

$$n = \frac{N}{1+N(e)^2} \quad (1)$$

n: sample, N: population, dan e: *error* [9]. A research flow chart is presented in FIGURE 1.

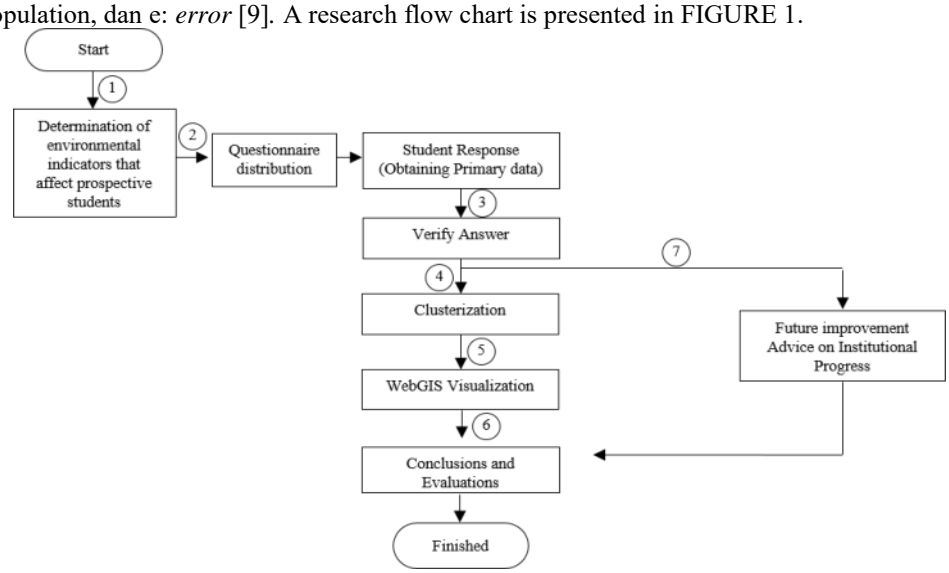


FIGURE 1. Research flow chart

The description of each stage includes (1) preparation in the form of questionnaires, determination of samples and indicators of research, (2) the process of distributing questionnaires is done in stages over three months, (3) data obtained from student responses. Primary data obtained was verified to minimize errors in the calculation of cluster analysis and other data analysis, (4) Cluster analysis with the aim of obtaining representative clusters/ groups according to the data held, (5) making WebGIS in each study program and overall, (6) conclusions and evaluation of the results of clustering and WebGIS analysis. The independent variable in the study consisted of five variables, namely: choice and advice from parents (X1), environmental recommendations such as siblings and neighbors (X2), successful people in the surrounding environment (X3), recommendations of teachers and schoolgirls when taking high school (X4), and the last variable is the influence of the elder brother during school

(X5). Control variables include the number of students and questionnaires given. While the terrestrial variables are cluster results obtained

Cluster Analysis

Clustering is a process used to group a set of objects in such a way that similar data objects are grouped, while the other objects are in another group [10]. The set-in clustering is known as the cluster name. Cluster objects produce high similarity levels in the same class and low similarity levels for different classes.

Cluster formation methods are generally categorized based on the type of data structure provided. In general, the cluster method is divided into two, namely Non-Hierarchical Clustering and Hierarchical Clustering. Based on the characteristics of the data in this study Hierarchical Clustering was used. This is because there is no information on the number of groups to be selected. In the Hierarchical Clustering, the grouping direction can be agglomerative (bottom-up), which means that from the n pieces of data to be clustered or divisive (top to down), it means that from one cluster to a cluster. From the Hierarchical Clustering Technique, we get a collection of sequential partitions, wherein the collection is obtained: (1) clusters that have individual points (these clusters are at the lowest level), and (2) a cluster in which there are points -points that all clusters have in it. This cluster is at the top level. The flow chart of agglomerative clustering analysis is as follows [11].

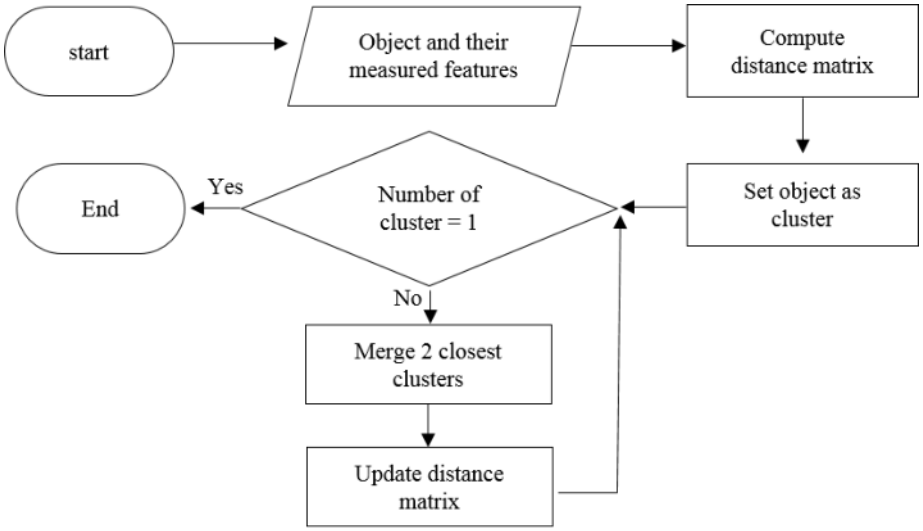


FIGURE 2. Flowchart of Agglomerative Clustering

Hierarchical Clustering steps [12] include: (1) identification of items based on the closest distance, (2) items that have been identified are combined in one cluster, (3) distance calculation of each cluster, (4) repetition until everything is connected. The assumptions used in cluster analysis include two things [13], namely: (1) a representative sample which means how much the sample represents the population and (2) the absence of multicollinearity using Pearson correlation between independent variables, if the correlation value exceeding 0.8, multicollinearity is said to occur. If there is multicollinearity, the Principal Component Analysis (PCA) method is used.

After assumptions are fulfilled, clustering can be done with Agglomerative Clustering. In general, the Agglomerative Clustering method is divided into 5 methods that are often used, namely: (1) Single Linkage, (2) Complete Linkage, (3) Average Linkage, (4) Ward's Method, and (5) Centroid Method. Single Linkage is based on the smallest distance. If two objects have a short distance, then the two objects will be combined in one cluster. As an illustration of the distance between clusters Cx and Cy, calculated by Eq

$$d(C_x, C_y) = \min_{\substack{x \in C_x \\ y \in C_y}} d(x, y) \quad (2)$$

Complete Linkage is based on the farthest distance calculated by Eq

$$d(C_x, C_y) = \max_{\substack{x \in C_x \\ y \in C_y}} d(x, y) \quad (3)$$

Average Linkage is similar to single linkage or complete linkage, but the criteria used are the average distance of all individuals in a cluster with the distance of all individuals in another cluster.

$$d(C_x, C_y) = \frac{1}{n_x n_y} \sum_{x=1}^{n_x} \sum_{y=1}^{n_y} d(x, y) \quad (4)$$

The concept of Ward's method is the formation of clusters based on loss of information due to the merging of objects into clusters. Merging 2 groups will be minimized Error Sum of Square Criterion (ESS)

$$ESS = \sum_{i=1}^{n}(X_i - \bar{X})^T(X_i - \bar{X}) \quad (5)$$

X_i: i-object vector, n : the number of objects in the cluster formed, $\bar{X}$: vector of the average of all objects [14]. While the centroid method identifies the distance between clusters is the distance between centroids. The centroid is the average of all objects in a cluster. The new centroid is calculated every time the object is combined so that every time there are additional members the centroid will change too.

After clustering, cluster validation was obtained. Validation aims to ensure that the cluster results obtained are representative of the population in general and can be used for other objects and are stable over some time. Although evaluation is generally not an integral part of cluster analysis, evaluation of the resulting clusters is quite important and better is part of cluster analysis [15].

Several cluster evaluation methods can be used, including Separation, Cohesion, and Cophenetic Correlation Coefficient. In this study Cophenetic correlation was used. Cophenetic Correlation was first introduced by Sokal [16] and is commonly used in numerical phenetic studies (classification of organisms based on similarities/similarities in morphology and observable other properties does not depend on the origin of the evolution of the organism). It is generally used as a measure of the degree of classification suitability and as a criterion for evaluating the efficiency of various grouping techniques [17].

WebGIS Visualization

Starting from Xerox Park Map Viewer which provides interactive website mapping sites in online information services [18]. WebGIS has been used in education since the mid-1990s such as GLOBE-based science education networks [19], KanCRN [20]. Journey North [21] uses WebGIS to collect, analyze and map data [22]. WebGIS also helps in terms of hardware limitations in mapping.

WebGIS is a combination of digital maps with geographic analysis, graphic mapping design, computer programming, and databases that are interconnected in one part of web design and web mapping [23]. The devices used in WebGIS include:

1. X Apache MySQL PHP Perl (XAMPP).
   Used as a server (localhost) to run web pages.
2. Sublime Text 3
   Used to edit data coding and various programming languages including PHP programming.
3. Microsoft Excel
   Used to calculate the spatial and non-participatory data of prospective students.
4. Browser Google Chrome dan Mozilla Firefox
   Used to access websites.
5. Quantum GIS
   Used to capture, store, analyze and display geospatial data [24]. The Quantum GIS used is QGIS 3.4 (Madeira).

The research data for making WebGIS is divided into two types, namely spatial data and nonspatial data. Spatial data in the form of provincial administrative data in Indonesia. While non-spatial data obtained from the results of clustering obtained. The flow chart of WebGIS is presented in FIGURE 3

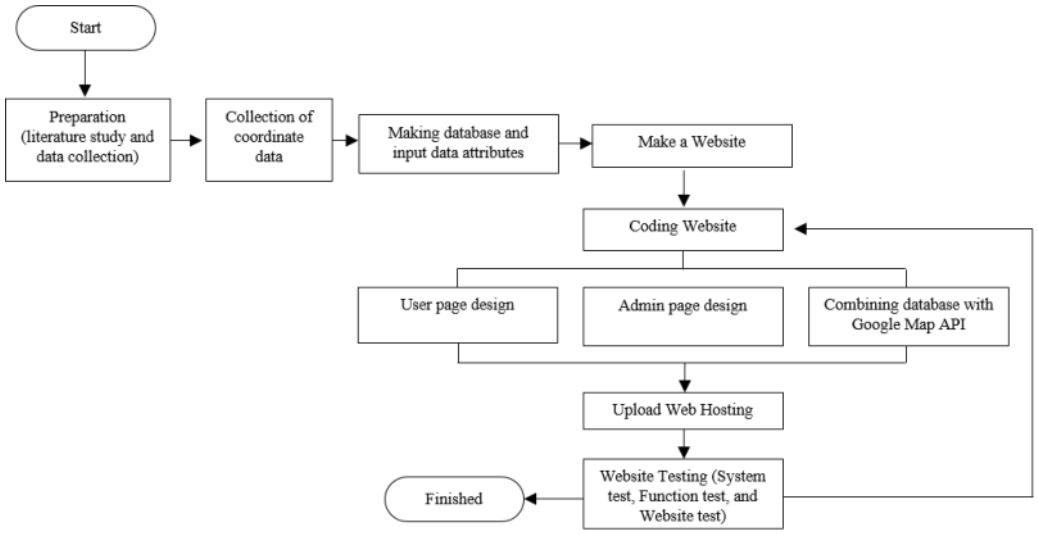


FIGURE 3. Flowchart of WebGIS Visualization

After the data and results of clustering are obtained, the next step is to create a database. The first time that is done before doing database design is making localhost using XAMPP. While to edit data coding used sublime text 3. The programming language used in making WebGIS is Javascript, Cascading Style Sheets (CSS), and Hypertext Markup Language (HTML). Javascript is used to make the web more dynamic and interactive directly with HTML. CSS is used to beautify web pages and control components in a web so that it is more structured and uniform. HTML is used to create a web page and display various information in the form of text and images on the website page.

RESULTS AND DISCUSSION

The data used came from each study program at FMIPA with the number of respondents taken by the proportion of total students in each study program. The next step is to verify the answers to minimize input errors or the resulting bias. Verification of answers included checking incomplete answer sheets, the number of students less than three in each province, or not according to the rules of filling, if one of the criteria is met, the data is not used.

Nevertheless, special chemistry education study programs elimination criteria are only applied if the filling is incomplete or errors in filling. Meanwhile, if there are less than 3 students from the province, the data is still used, this is because the Chemistry education study program is included in the new study program category and the number of students is still small for each province (an average of 4 students in each province). More complete can be seen in TABLE 1.

TABLE 1. Descriptive Statistics of prospective student data

Department	Provinces that have minimum students	Average	Number of Provinces	Maximum (Province)
Statistics	(Bali, DKI, Kalimantan Utara, Maluku, NTT, Papua, Sumatra Selatan, Sumatra Utara)	6.67 ≈ 7	28	43 (Jawa Tengah)
Chemistry	(Kalimantan Tengah, Maluku, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Utara, Sumatra Utara)	8.3 ≈ 8	23	35 (DIY)
Pharmacy	(Maluku, Maluku Utara, Papua Barat, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Utara, Sulawesi Barat, Sumatra Utara)	8.687 ≈ 9	32	55 (DIY)
Chemical Analyst	(Aceh, Kalimantan Tengah, Maluku Utara, Sulawesi Tengah, Sulawesi Tenggara)	9.65 ≈ 10	23	60 (Jawa Barat)
Chemistry Education	(Banten, Bengkulu, Kalimantan Barat, Kalimantan Timur, Kalimantan Utara, Maluku Utara, NTB, NTT, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tenggara)	4.17 ≈ 4	23	31 (Jawa Tengah)

The second column shows the provinces where the number of students is the least and the number of students from the province is the minimum number of students. From the third and fourth tables, it appears that the information is quite interesting, namely the Statistics study program has not too much average, but the distribution is quite extensive. While Pharmacy has a high average and widespread. The last column shows the province with the most prospective students, for example, the province's Chemical Education study program with the most prospective students in Central Java with a total of 31 students. Then for the Multicollinearity test, a correlation value is used. Obtained value for all study programs there is no multicollinearity (no correlation value exceeds

0.8). As an illustration, the following are the results of multicollinearity tests of Statistics and Chemistry study programs

Data: multikol

	row.names	X1	X2	X3	X4	X5
1	X1	1.00000000	0.21962966	0.08971364	-0.1315961	0.43999910
2	X2	0.21962966	1.00000000	0.65518054	0.3186935	-0.07188321
3	X3	0.08971364	0.65518054	1.00000000	0.5573097	-0.07031684
4	X4	-0.13159606	0.31869347	0.55730968	1.0000000	0.26594711
5	X5	0.43999910	-0.07188321	-0.07031684	0.2659471	1.00000000

(a)

Data: multikol

	row.names	X1	X2	X3	X4	X5
1	X1	1.00000000	-0.4554931	0.19744008	0.07621502	0.04280023
2	X2	-0.45549313	1.0000000	0.17542214	-0.30155203	-0.28877758
3	X3	0.19744008	0.1754221	1.00000000	0.36507984	0.04270022
4	X4	0.07621502	-0.3015520	0.36507984	1.00000000	0.73325922
5	X5	0.04280023	-0.2887776	0.04270022	0.73325922	1.00000000

(b)

FIGURE 4. (a) Multicollinearity test of Statistics department, (b) Multicollinearity test of Chemistry department

After verification of the answers and the previous provisions of only 60-70%, can be seen in TABLE 2.

TABLE 2. Provincial distribution in the clustering process

Prodi	Questionnaire Data		Used Data	
	Number of Provinces	Number of Students	Number of Provinces	Number of Students
Statistics	28	187	15	169
Chemistry	24	197	13	177
Pharmacy	32	278	19	259
Chemical Analyst	23	222	11	203
Chemistry Education	23	97	23	97

Then from the data that has been verified, clustering is done with Agglomerative Clustering. The five methods of Agglomerative Clustering are used, namely Single Linkage, Complete Linkage, Average Linkage, Ward's Method, and Centroid Method. The five methods will be compared to produce the best clusters. The size of the distance used is Euclidean Distance. The distance between objects x and y is calculated based on Eq

$$d_n(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (6)$$

$x : (x_1, x_2, \cdots, x_n)$ and $: (y_1, y_2, \cdots, y_n)$ are two data objects with n attributes [25].

From the results of clustering, the number of clusters from each study program was different between study programs. This is because the characteristics of prospective students also vary. Obtained by 4 clusters for a Statistics study program, 5 clusters for Chemistry study program, 4 Pharmacy study cluster, 2 Chemistry Analyst study program clusters, and 2 clusters for Chemistry Education study program. The visualization of cluster results is presented in a dendrogram. Dendrograms are tree diagrams as a form of graphical representation of the hierarchical cluster [26].

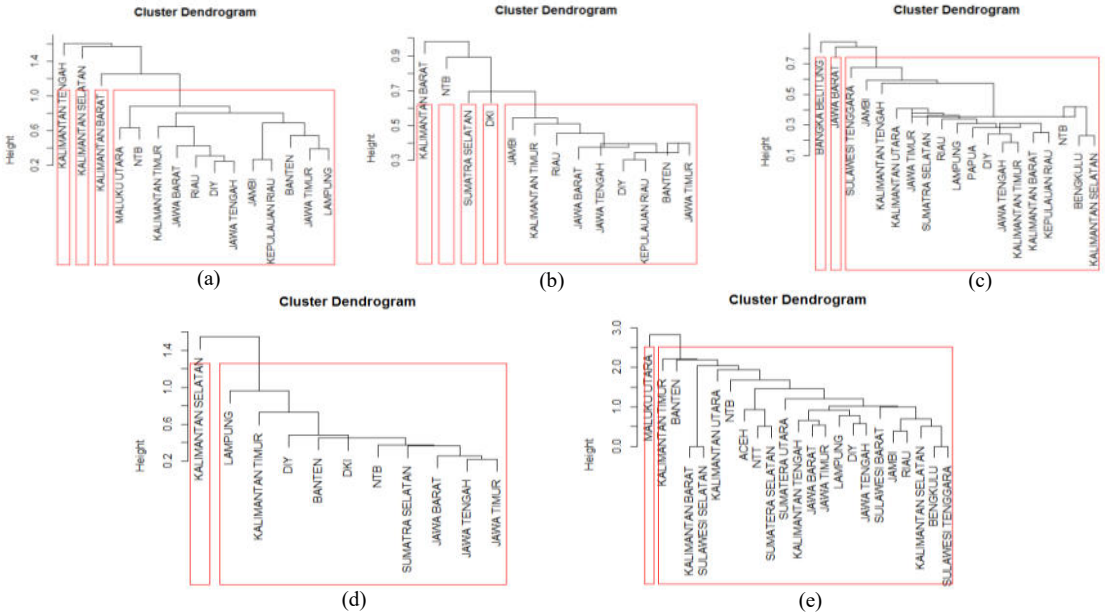


FIGURE 5 . (a) Statistics, (b) Chemistry, (c) Pharmacy, (d) Chemical Analyst, (e) Chemistry Education

Then from each study program, the value of the Cophenetic Correlation Coefficient obtained from each method was obtained. The cophenetic correlation coefficient is the correlation between the distance of each pair of objects calculated based on the concept of distance and the merging distance that is generated from the merging method used. Indicators of this evaluation are presented with indicators of cophenetic values. Cophenetic value is a

quantitative measure used to evaluate the quality of the Hierarchical Clustering method, calculated using the Cophenet function according to Equations [27].

$$c = \frac{\sum_{i<j}(x(i,j)-x)(t(i,j)-t)}{\sqrt{[\sum_{i<j}(x(i,j)-x)^2][\sum_{i<j}(t(i,j)-t)^2]}} \quad (7)$$

The clustering is said to be valid if there is a strong correlation between the two distances. It can also be said, a Hierarchical Clustering result is said to be significant if the Cophenetic value is close to one (1). The Cophenetic Correlation Coefficient value is displayed in TABLE 3.

TABLE 3. Value of Cophenetic Correlation Coefficient

Departments	Average Linkage	Centroid Linkage	Complete Linkage	Single Linkage	Ward's Method
Statistics	0.8988396	0.8847553	0.8689974	0.8739437	0.6083208
Chemistry	0.8186653	0.818986	0.7932464	0.7564301	0.6642111
Pharmacy	0.7813897	0.7822756	0.6124972	0.7399113	0.4861627
Chemical Analyst	0.9379985	0.9398861	0.9135999	0.923701	0.812555
Chemistry Education	0.8588379	0.8391886	0.5129312	0.8268961	0.4474618

From the TABLE the Cophenetic Correlation Coefficient values obtained from each study program for the five methods. Each study program has a different clustering method. Statistics Study Program obtained the largest Cophenetic value for the Average Linkage method which is equal to 0.8988396 so the clustering used is the Average Linkage method. The analogy of the principle is obtained by the method used for each study program.

After clustering, the distribution of provincial clusters in which the surrounding environment affects prospective students to choose study programs at FMIPA UII. For example, from cluster results obtained 5 clusters (from cluster I to cluster V), then sorted by how much influence the environment on prospective students choose in the study program. Cluster I states that environmental factors have no effect and sequentially up to cluster V which states that the environmental factors have the greatest effect (major effect). Similarly, if the number of clusters is only 3 and 4, while if the number of clusters is only divided into 2 clusters, it can be said to be a small and large effect. The image is presented in TABLE 4.

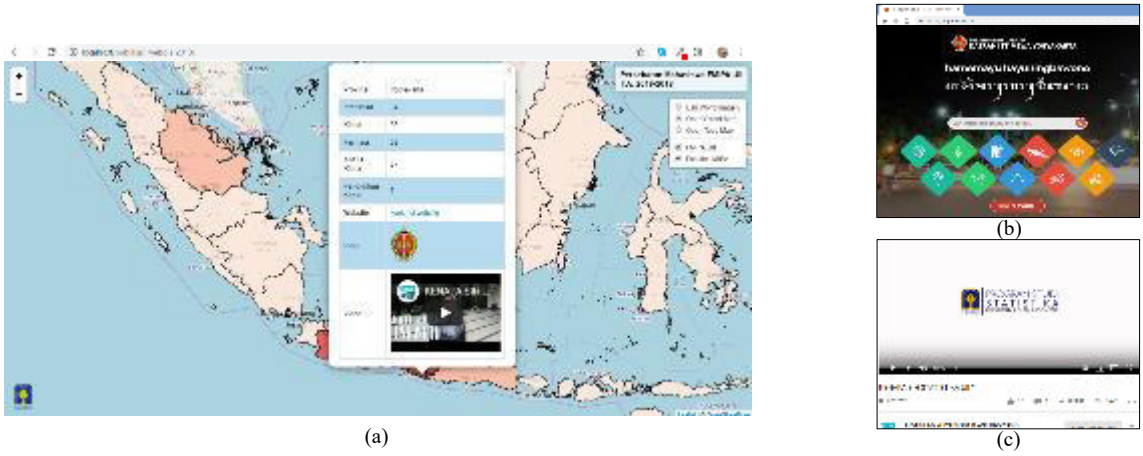
Then from the results of clustering, data visualization is done using WebGIS. Data visualization is the compilation of data in several more systematic forms including attributes and variables used for information delivery [28]. As an initial description, the following is a visualization of the number of prospective students in the Faculty of Mathematics and Natural Sciences.

From the basic visualization, WebGIS (Figure 6) can be filled in attributes from each province, such as the number of students, descriptions of each location, and additional images to draw the visualization given. Furthermore, the addresses of links and profile videos of each student from the province can be given. This visualization can be used as an interactive and dynamic media as a promotion for prospective students in subsequent periods. After obtaining cluster results for each province, processed data in the making of WebGIS visualization between real data and cluster results are presented in FIGURE 7 (for example used visualization cluster WebGIS Statistics department).

The darker the color produced shows the greater the influence of environmental factors on the motivation of prospective students choosing statistics study program, and vice versa. Based on FIGURE 7, an interesting result is obtained, namely clustering data results when compared with the original data, so it cannot be concluded as a whole that environmental factors influence prospective students to choose a statistics study program. As an example of East Kalimantan province, the cluster results of East Kalimantan province are included in cluster 4 (which means that environmental influences are very strong) but the real data of the number of prospective students are not comparable to the results of small clustering (if clustered into cluster 1). Unlike the case with the province of Central Java, the results of clustering have a value comparable to the number of students, which is in cluster 4 and the number of students is in a very large category (cluster 4).

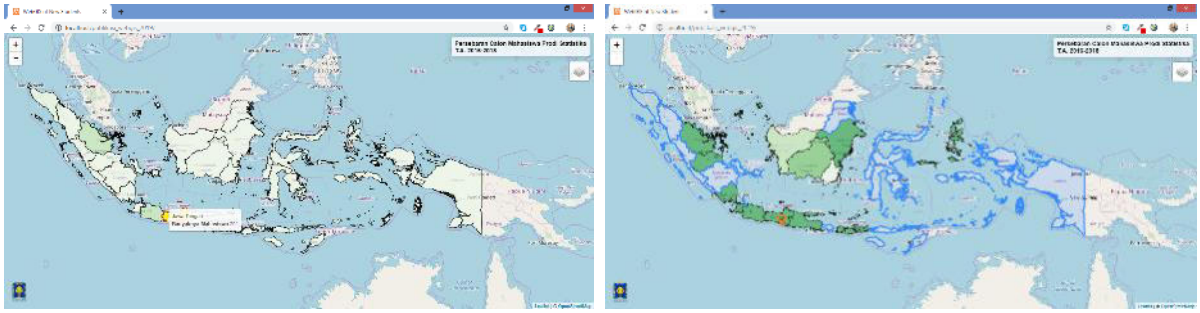
TABLE 4. Provincial distribution in the clustering process

Department	Best Clustering Method	Number of Clusters	Cluster				
			I	II	III	IV	V
Statistics	Average	4	Kalimantan Selatan	Kalimantan Barat	Kalimantan Tengah	Banten, DIY, Jambi, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Timur, Kepulauan Riau, Lampung, Maluku Utara, NTB, dan Riau	-
Chemistry	Centroid	5	NTB	Sumatra Selatan	DKI	Banten, DIY, Jambi, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Timur, Kepulauan Riau, Riau	Kalimantan Barat
Pharmacy	Centroid	3	Bangka Belitung	Jawa Barat	Bengkulu, DIY, Jambi, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Kepulauan Riau, Lampung, NTB, Papua, Riau, Sulawesi Tenggara, Sulawesi Selatan	–	–
Chemical Analyst	Centroid	2	Banten, DIY, DKI, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Timur, Lampung, NTB, Sumatra Selatan	Kalimantan Selatan	–	–	–
Chemistry Education	Average	2	Aceh, Banten, Bengkulu, DIY, Jambi, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Lampung, NTB, NTT, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Utara	Maluku Utara	–	–	–



(a) (b) (c)

FIGURE 6. (a) Basic Visualization of WebGIS, (b) website link, (c) video



(a) (b)

FIGURE 7. (a) WebGIS real data for prospective students in Statistics, (b) WebGIS from Statistics department clusters

Based on this, if the number of real student candidates from an area falls into a very large category (4) but after clustering is entered in clusters 1 or 2, then the influence factor of the surrounding environment is not too significant. It can also be said that the prospective student chooses the study program due to other factors, such as graduated alumni, motivation in themselves, or even job opportunities from the study program concerned. If in conditions like this, study programs can optimize such as strengthening the dissemination of information relating to prospects and dissemination of promotions. Meanwhile, if the opposite conditions apply (the number of small students (1) but are in clusters 3 or 4), then communication with alumni can be strengthened as well as strengthening media promotion in the area. From this evaluation, it is expected that optimal results can be obtained not only quantity but also quality.

CONCLUSIONS

Based on the results of clustering, the number of different clusters was obtained. Statistics Study Program is divided into 4 clusters, Chemistry 5 clusters, Pharmacy 3 clusters, Chemical Analyst and Chemistry Education 2 clusters. The respective clusters in the form of levels from cluster I are very ineffective until V which means that environmental factors greatly affect the prospective students to choose the study program. Then WebGIS was used to simplify the visualization of the clusters obtained. From the results of WebGIS visualization, it was found that it cannot be said that prospective students in the smallest cluster have a small number of students. This is because there are still other factors such as alumni that have been produced, the motivation of the prospective students, and even future employment opportunities. This is a challenge for each study program in future policymaking. With optimal policymaking, optimal results will also be obtained not only in terms of quantity but also quality.

REFERENCES

[1] M. J. Marsden, "Evaluation: Towards a definition and statement of Purpose," *Australian Journal of Educational Technology,* pp. 31-38, 1991.

[2] S. V. Reddy, S. Subash, and P. Prakash, "Visualization of Big Data: Its tools and challenges," *International Journal of Applied Engineering Research 9(18),* pp. 5277-5290, 2014.

[3] P. Novianti, D. Setyorini and U. Rafflesia, "K-Means cluster analysis in earthquake epicenter clustering," *International Journal of Advances in Intelligent Informatics Vol 3, No.2,* pp. 81-89, 2017.

[4] Z. Zhang, F. Murtagh, S. V. Poucke, S. Lin and P. Lan, "Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R," *Annals of Translational Medicine, Vol 5, No 4. doi: 10.21037/atm.2017.02.05,* 2017.

[5] S. Rana and R. Garg, "Application of Hierarchical Clustering Algortihm to Evaluate Students Performance of an Institute," in *Second International Conference on Computational Intelligence & Communication Technology*, India, 2016.

[6] Asroni, H. Fitri and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Teknika,* vol. 21, no. 1, pp. 60-64, 2018.

[7] A. Bendib, D. Hadda and K. Mahdi, "Application of WebGIS in the development of an interactive interface for urban management in Batna City," *Journal of Engineering and Technology Research,* pp. 13-20, 2016.

[8] R. Lathrop, L. Auermuller, J. Trimble and J. Bognar, "The Application of WebGIS Tools for Visualizing Coastal Flooding Vulnerability and Planning for Resiliency: The New Jersey Experience," *ISPRS International Journal of,* pp. 408-429, 2014.

[9] T. Yamane, Statistics, An Introductory Analysis, 2nd Ed, New York: Harper and Row, 1967.

[10] S. S. Ghuman, "Clustering Techniques-A Review," *International Journal of Computer Science and Mobile Computing,* pp. 524-530, 2016.

[11] P. Verma and M. Mahajan, "Hierarchical Clustering Approach for Modeling of Reusability of Function Oriented Software Components," 2011.

[12] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Amsterdam: Elsevier, 2012.

[13] W. Alwi and M. Hasrul, "Analisis Klaster untuk Pengelompokan Kabupaten/Kota di Propinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat," *Journal MSA Vol 6 No.1 Ed. Jan-Jun 2018,* pp. 35-42, 2018.

[14] R. A. Johnson and D. W. Wichern, Applied Multivariate Statistical Analysis Sixth Edition, New Jersey: Pearson Prentice Hall, 2007.

[15] P.-N. Tan, M. Steinbach, A. Karpatne and V. Kumar, Introduction to Data Mining, Boston: Pearson Education, 2006.

[16] R. R. Sokal and F. J. Rohlf, "The Comparison of Dendrograms by Objective Methods," *Taxon Vol.11 No.2,* pp. 33-40, 1962.

[17] J. S. Farris, "On the Cophenetic Correlation Coefficient," *Systematic Biology, Volume 18, Issue 3,* pp. 279-285, 1969.

[18] Xerox Park Map Viewer, "An Interactive Map Viewer," 2017. [Online]. Available: https://web.archive.org/web/20060916133035/http://www2.parc.com/istl/projects/www94/mapviewer.html.

[19] GLOBE, "The GLOBE Program," 1994. [Online]. Available: https://www.globe.gov/about/history.

[20] KanCRN, 1996. [Online]. Available: http://www.pathfinderscience.net/about.

[21] Journey North, "Journey North.," 1997. [Online]. Available: https://journeynorth.org/maps/.

[22] T. R. Baker, "WebGIS in Education," in *Geospatial Technologies and Geography Education in a Changing World*, Tokyo, Springer, 2015, pp. 105-115.

[23] N. Qolis and A. Fariza, "Pemetaan dan Analisa Sebaran Sekolah untuk Peningkatan Layanan Pendidikan di Kabupaten Kediri dengan GIS," EEPIS Final Project, Surabaya, 2010.

[24] K.-t. Chang, Introduction to Geographic Information Systems 9th Edition, New York: McGraw-Hill Education, 2019.

[25] M. D'Agostino and V. Dardanoni, "What's so Special About Euclidean Distance? A Characterization With Applications to Mobility and Spatial Voting," *Social Choice and Welfare,* p. 211–233, 2009.

[26] F. A. Espinoza, J. M. Oliver, B. S. Wilson, and S. L. Steinberg, "Using Hierarchical Clustering and Dendrograms to Quantify the Clustering of Membrane Proteins," *Bulletin of Mathematical Biology,* 13 July 2011.

[27] Mathworks statistics toolbox, "Mathworks," 2012. [Online].

[28] M. Khan and S. S. Khan, "Data and Information Visualization Methods, and Interactive Mechanisms: A Survey," *International Journal of Computer Applications,* pp. 1-14, 2011.