Jurnal Pijar MIPA
https://jurnalfkip.unram.ac.id/index.php/JPM

ISSN 1907-1744 (Print)
ISSN 2460-1500 (Online)

Determining Zoning of Areas Affected by Flood Disasters in Medan City Using Silhouette Coefficient and Davies Bouldin Index Analysis

Novi Eliza Poetri*, Ismail Husein, Sajaratud Dur

Mathematic Study Program, Faculty of Science and Technology, North Sumatra State Islamic University, North Sumatra, Indonesia
*E-mail: novielizapoetri@gmail.com

Received: April 9, 2024. Accepted: April 28, 2024. Published: May 30, 2024

Abstract: Several sub-districts are cities that are pretty or even very vulnerable to flood disasters. Therefore, the Government must observe which areas are highly prone to flooding to anticipate safety precautions. One of the relevant methods for handling this case is k-means clustering using the Silhouette Coefficient and Davies-Bouldin Index evaluation. This research uses a quantitative approach using the Silhouette Coefficient and Davies Bouldin Index Analysis method. Based on the research that has been carried out, the results can be obtained that cluster 1 consists of the sub-districts of Medan Polonia, Medan Johor, Medan Selayang, Medan Baru, Medan Tuntungan and Medan Barat. Cluster 2 only has one member, namely the Medan Maimun sub-district; Cluster 3 consists of Medan Sunggal and Medan Belawan sub-districts, and Cluster 4 consists of Medan Marelan, Medan Helvetia and Medan Timur sub-districts. The results of the clustering evaluation that was carried out obtained a Silhouette Coefficient value of 0.505539942 and a Davies-Bouldin Index value of 0.30055. This means that the grouping carried out in this research is accurate.

Keywords: Clustering; Davies Bouldin Indeks; K-Means; Silhouette.

Introduction

Indonesia is a country that geographically has a tropical climate with relatively high humidity in almost all regions in Indonesia. This is caused by high rainfall every time the rainy season occurs [1]. Apart from that, high rainfall can also cause adverse effects such as floods and landslides. Flood disasters in Indonesia can be said to be seasonal disasters because every time the rainy season arrives, floods hit several areas in Indonesia. Floods occur due to surface water runoff that overflows, and the volume exceeds the drainage capacity of such drainage or water bodies [2]. The five factors that cause floods in Indonesia include rain factors, factors that damage the retention of river watersheds (DAS), factors that cause errors in planning the development of river channels, factors that cause river shallowing, and factors that cause errors in regional planning and construction of facilities and infrastructure [3]

Flooding is an event where land that is usually dry becomes inundated with water, which is caused by high rainfall and the topographic conditions of the area that is beautiful to the point of being submerged and can also be caused by the poor ability of infiltration by the soil, resulting in the soil being no longer able to collect and absorb debris Water [2]. Based [3] believes that floods are the most frequently occurring natural disasters, both in terms of their intensity in some regions of Mauipu and the number of occurrence locations in our knowledge is around 40% of other natural disasters.

One of the flood-prone areas in Suimateira Uitara is Meidan City, with an industrial population of around 2,460,858 people and an area of 26,510 ha or 3.6% of the total area of Suimateira Uitara Province [4]. Floods in Meidan City occur almost 10-12 times/year, which is influenced by the occurrence of Suingai Deili and Beilawan in the Huilui area. Kabuipatein Karo, Kabuipatein Deili Seirdang, and Meidan City are areas prone to flood disasters caused by the influx of water from the Huilui area and the deplorable drainage conditions in the city [5].

Over a century, the recorded flood disasters that have hit Indonesia, if viewed from the frequency intensity, have reached 108 times, or if analyzed, it has reached 33.3% of all natural disaster events that have ever hit Indonesia. In this case, the impact of significant disasters struck Indonesia and reached 324 events. The existence of flood events is influenced by two factors, namely natural and anthropogenic. It is known that the main factor that caused this disaster to arise was rainfall with high intensity and long-term rain; the factor that caused this disaster was the weak awareness of land use management (landuisei) in the affected area. This statement is confirmed by the opinion [6], which states that the factors causing flooding are high rainfall intensity and poor monitoring of land conservation.

The problem of flooding in the city of Meidan has entered the chronic disease category and has even become a traditional tradition. Many of the various efforts have been carried out intuitively without considering the large amount spent on multiple projects. To date, the flood has affected 2.1 million people in Meidan. This is because the current floods do not depend on the rain that occurs in the huilui suingai Deili alone; the rain in the city of Meidan can cause the Meidan people to be overwhelmed because the water pools everywhere. This also means that water areas that are densely populated by people will become prone to flood

How to Cite:
Eliza, N., Husein, I., & Dur, S. (2024). Determining Zoning of Areas Affected by Flood Disasters in Medan City Using Silhouette Coefficient and Davies Bouldin Index Analysis . *Jurnal Pijar Mipa*, *19*(3), 558–563. https://doi.org/10.29303/jpm.v19i3.6707

protection, especially as heavy rainfall in the Suingai-Suingai section of the Huilui River crosses the city of Meidan [7]. For this reason, the City of Meidan is taking preventative and suitable balancing measures in all areas, especially in mitigating floods that occur during the rainy season and even become a regular occurrence for many people. Therefore, this activity is carried out by groups of areas prone to other natural disasters.

The selection of this method is based on research that has been carried out [8], which shows that one of the research methods that is simple and has a high optimal value is the K-Meians method. Cluster analysis requires that the coefficient coefficient and Davieis bouildin index be applied because the researcher wants to know whether it is clustering by choosing the correct and optimal k and a robust data structure. This is also done in research [9], where silhouette analysis is carried out in intuitive inquiry to see the similarity of objects from intuitive data. On the other hand, analyzing the structural strength of data in experiments with several clusters is necessary to understand whether data is transferred from one cluster to another.

In data mining applications, clustering plays a significant role, for example, in scientific data exploration, web analysis, text mining, and spatial database applications. In this research, clustering is used to identify disaster-prone areas. It will be known which areas are highly affected by natural disasters such as tornadoes [10]. The clustering process begins by identifying the data to be grouped. Then, the distance between each cluster center and each piece of data is calculated. The Euclidean formula is used to find out this distance [11].

Silhouieittei coefficient is an optimal intuitive evaluation method or the occurrence of a cluisteir that has been intuitively derived from the cluisteiring process [12]. Silhouieittei coefficient shows the results of the visual quality of objects in each cluisteir, giving information according to the number of cluisteirs in the set data. The intuition of each object is denoted by the clue where it originates [13]. This method combines the separation and cohesion methods [14].

Bouildin developed a cluisteir evaluation method named by the researcher "Davieis-Bouildin Index (DBI)." DBI is one of the internal cluisteir evaluations, where whether or not the cluisteir results are good can be assessed from coherence and separation [15]. Cohesion is the extent to which the distance between data is intuitively implied and the extent of the similarity of the data on the same cluster [16]. Meanwhile, the intuitive distance values are separated between different data sources. Intuition requires DBI to maximize the inter-cluster distance between clusters cj and ci and simultaneously try to intuitively minimize the distance between points within a cluster [17]. If the intra-cluster distance is minimal, each object in the cluster will be seen more clearly [18].

The choice of cluster analysis is also based on research conducted by [19], which shows that you can find the optimal number of clusters from several cluster analyses. In his research, he used the C-Index, Davies Bouldin Index, McClain Rao Index, Silhouette Index, and Connectivity Index to find that the number of clusters formed was 2 clusters and 4 clusters. Using cluster variance, it was found that the best cluster was to use 4 clusters because it had the smallest Sw and Sb ratio values.

In this research, a representation of the cluster results was also carried out using GIS coloring, as explained by [20] in his study on earthquake-prone zoning in the city of Malang that natural disaster-prone zoning is essential because it is one of the tries to minimize losses of property and life caused by natural disasters. To make it easier for the Government to identify areas prone to flood disasters after the correct clusters have been formed, the regions are colored into a map of areas prone to flood disasters using the Geographic Information System (GIS).

Based on the Silhouette Coefficient and Davies Bouldin Index analysis, it can be identified which areas (cities or sub-districts) have a high intensity of being affected by flood disasters. This is being pursued because it is one of the applications of disaster mitigation, meaning that research is being carried out to become a reference for the Government in preparing for flood disasters.

Research Methods

This quantitative research uses complete data from the National Research and Development Agency (BNPB) research carried out in 2018.

Table 1. Variable Operational Definition [2]

No.	Variable	Denition	Unity	Mark
1.	Count Flood Disaster Events	Number of flood disaster events	Count the Events	53
2.	Disaster Victims	Soul	Number of Victims	13504
3.	Home damage	Damage The building collapsed	Number of Houses	17
4.	Uinit's confusion (Facility)	Uinit's confusion (treasure)	Number of units damaged	11

The analysis steps in this research are to collect various scientific materials such as material obtained in libraries and academic materials, and national and international research journals. Then, we want to compile data from the National Peinang Guilangan Beincana Agency (BNPB) or the Regional Peinang Guilangan Beincana Agency (BPBD) website. Carry out data processing. Data processing is carried out in the following stages: clustering is done using the K-Meians algorithm once the data is obtained. Where it can be known which groups in each area are prone to flood disasters. Once the cluisteir is intuitive, the next step is to conduct a cluisteir validation test. This is done intuitively to determine the strength of the data structure from the number of clues. In research, we will need several methods to evaluate the clues, including Intuiting the number of cluisteir using the Silhouieittei coefficient. Davieis bouildin Indeix and the final result of the research is a characteristic analysis of the zoning of areas prone to flood disasters in Mei City and wanting the K-Means algorithm with analysis of silhouiette coefficient and Davies Bouldin index. Next, intuit the starting point of the centroid (central point), then calculate the distance from the object to the center. Group objects based on minimum distance to evaluate whether they need to be moved; literacy will continue or continue if there are no objects that move, but on the contrary, if something moves, it will return to the initial step and generate conclusions and suggestions.

Results and Discussion

Data were experimented with on cluster 1, cluster 2, cluster 3, and 4 with the aim of intuitive research to find out which cluster was more accurate. The first step that will be taken in gathering K-Meians is to take 12 random samples of data, namely in the districts of Meidan Polonia, Meidan Johor, Meidan Seilayang, Meidan Barui, Meidan Tuintuingan, Meidan Barat, Meidan Maimuin, Meidan Suinggal, Meidan Beilawan, Meidan Mareilan, Meidan Heilveitia and Medan Timur.

Table 2. Ceintroid iteration 1

Subdistrict	a	b	c	d	e	f	g	h	i	J	k
Medan Johor	0	61	0	80	62	7	41	0	0	0	0
Medan Maimun	0	71	0	57	70	6	42	0	0	0	0
Medan Sunggal	0	81	0	87	80	9	57	0	0	0	0
Medan Timur	0	90	0	80	90	9	66	0	0	0	0

Data from 12 sub-districts will be used as a sample using random sampling as a sequence.

After intuiting the value used as the initial centroid, a process will be carried out to intuit the distance of each data to the centroid based on Eiuiclideian Distance. Eiuiclideian Distance from Medan Polonia District (1st data) to Medan Johor District (1st data) with equality.

$$d_{(1,1)} = \sqrt{\sum_{i=0}^{n} (xi - yi)^2} \qquad 1)$$

$$= \sqrt{\begin{array}{c}(0-0)^2 + (55-0)^2 + (0-0)^2 + (56-80)^2 + (50-62)^2 \\ +(9-7)^2 + \\ (25-41)^2 + (0-0)^2 + \end{array}}$$

$$= \sqrt{(0-0)^2 + (0-0)^2 + (0-0)^2}$$

= 31,87475

Eiuiclideian Distances behave intuitively as a sample of data with centroidal illusions. If the results are successful, the smallest or minimum Eiuclideian Distance will be declared a clue result in iteration 1. Follow the results of calculating the Euclidean Distance over the sample data.

Table 3. Euclidean Distance Analysis 1

Subdistrict	C1	C2	C3	C4	Result
Medan Polonia	31.87475	30.9030742	59.6741149	71.28814768	C2
Medan Johor	0	26.3628526	32.1403173	47.47630988	C1
Medan Selayang	42.1307488	26.3438797	62.1449917	69.02897942	C2
Medan Baru	27.6586333	117473401	38.9358446	45.37620522	C2
Medan Maimun	26.3628526	0	36.5239647	43.30127019	C2
Medan Tuntungan	26.6458251	25.6709953	30.5122926	41.01219331	C2
Medan Sunggal	32.1403173	36.5239647	0	17.63519209	C3
Medan Barat	28.7576076	29.0860791	41.3521462	51.29327441	C1
Medan Belawan	31.144823	26.2868788	16.0934769	22.8035085	C3
Medan Marelan	40.6448028	26.7020598	24.3926218	21.02379604	C4
Medan Helvetia	38.9615194	28.7923601	19.8746069	15.93737745	C4
Medan Timur	47.4763098	43.3012701	17.6351920	0	C4

Based on the table above, it can be seen that in the Meidan Polonia sub-district, the Euclidean Distance, the minimum distance or the independent distance, is 30.90307428, which is at C2, so it can be stated that the Meidan Polonia sub-district in iteration 1 is in the cluisteir 1. This can also be interpreted as other data; the closest distance or minimum value indicates the cluisteir group.The next step is to calculate the average value of the data based on cluisteir. This table displays the results of grouping data based on cluster (Table 4).

The data above shows that those included in the Cluisteir 1 group are the districts of Meidan Johor and Meidan Barat. And so on until Ceintroid Cluisteir 2, 3, and 4. Based on the calculations' results from each cluster's classification, an average value has been obtained, which will be formed as a new Centroid, as shown in table 5.

After achieving a new centroid, the next step will be to calculate (Table 6). Based on the table above, it can be seen that member cluster 1 (Table 7) and so on until the centroid clusters 2, 3, and 4 in iteration 2.

From the results of the second iterative calculation, it was found that several sub-districts experienced changes, namely cluisteir displacements, so that a return Eiuclideian Distance calculation could be carried out. The thing used as the latest intuitive centroid for the 3rd iteration is the average value of each clue in the 2nd iteration. The newest centroid from the results of the second iteration of calculations can be seen in the following table 8. Once a new centroid is obtained, the next step is to carry out the third Eiuclideian Distance iteration.

The calculation results can be seen in the following table 9. Based on the table 9, it can be seen that there is no longer a cluisteir that has moved to another cluisteir, or in other words, the new cluisteir is the same as the previous cluisteir. That intuition, the iteration, has come to an end. So it can be concluded that cluisteir 1 consists of keic. Meidan Polonia, Meidan Johor, Meidan Seilayang, Meidan Barui, Meidan Tuintuingan and Meidan Barat. Cluisteir 2 is keic. Meidan Maimuin. Cluisteir 3 is keic. Meidan Suinggal and Meidan Beilawan. Cluisteir 4 namely i, keic. Meidan Mareilan, Meidan Heilveitia and Meidan Timuir. This table shows the results of the 1st conclusion in the 3rd Eiuiclideian Distancei iteration calculation.

Table 4. Centroid Cluster 1 iteration 1

Subdistrict	a	b	c	d	e	f	g	h	i	j	k
Medan Johor	0	61	0	80	62	7	41	0	0	0	0
Medan Barat	0	56	0	71	86	6	29	0	0	0	0
Rata-Rata	0	58.5	0	75.5	74	6.5	35	0	0	0	0

Table 5. Iteration intuitive ceintroid 2

	a	b	c	d	e	f	G	h	i	j	k
C1	0	58.5	0	75.5	74	6.5	35	0	0	0	0
C2	0	163.5	0	141.5	162	18.5	89.5	0	0	0	0
C3	0	82.5	0	81.5	81	8	51.5	0	0	0	0
C4	0	86	0	66.5	84.5	8.5	56.5	0	0	0	0

Table 6. Euclidean Distance Analysis 2

Subdistrict	C1	C2	C3	C4	Iteration 1	Iteration 2
Medan Polonia	32.78337994	189.4122488	55.4143483	57.04384279	C2	C1
Medan Johor	14..3788	163.624	30.60	39.44617	C1	C1
Medan Selayang	39.2141	273.495	56.82	52.37366	C2	C1
Medan Baru	29.1333	167.370	35.11	30.75711	C2	C1
Medan Maimun	163.012	23.7434	30.78	27.23968	C2	C2
Medan Tuntungan	19.5128	149.803	24.07	31.09662	C2	C1
Medan Sunggal	34.1284	132.842	8.046	21.58703	C3	C3
Medan Barat	14.3788	161.613	36.71	41.04875	C1	C1
Medan Belawan	28.9093	137.967	8.046	14.59452	C3	C3
Medan Marelan	37.6264	137.277	18.10	5.385165	C4	C4
Medan Helvetia	38.3373	137.353	16.72	5.385165	C4	C4
Medan Timur	47.2837	122.519	18.72	17.86057	C4	C4

Table 7. ceintroid cluisteir 1 iteration 2

Subdistrict	a	b	c	d	e	f	G	h	i	j	k
Medan Polonia	0	55	0	56	50	9	25	0	0	0	0
Medan Johor	0	61	0	80	62	7	41	0	0	0	0
Medan Selayang	0	58	0	39	60	7	32	0	0	0	0
Medan Baru	0	67	0	55	64	7	51	0	0	0	0
Medan Tuntungan	0	76	0	76	80	8	29	0	0	0	0
Medan Barat	0	56	0	71	86	6	29	0	0	0	0
Rata-rata	0	62.16	0	62.83	67	7.33	34.5	0	0	0	0

Table 8. Updated centroid from 2nd iteration calculation

	a	b	C	d	e	f	g	H	i	j	k
C1	0	62.16	0	62.83	67	7.333	3.5	0	0	0	0
C2	0	71	0	57	70	6	42	0	0	0	0
C3	0	82.5	0	81.5	81	8	56.5	0	0	0	0
C4	0	87.33	0	71	86.33	8.66	59.66	0	0	0	0

Table 9. Results of Eiuclideian Distance Analysis 3

Sundistrict	C1	C2	C3	C4	Iteration 1	Iteration 2	Iteration 3
Medan Polonia	21.91080411	30.90307	55.41435	61.58282	C2	C1	C1
Medan Johor	19.06349361	26.36285	30.60637	41.44608	C1	C1	C1
Medan Selayang	25.31303752	26.34388	56.82209	57.84558	C2	C1	C1
Medan Baru	11.74734	19.13330208	35.11054	35.3003	C2	C1	C1
Medan Maimun	13.38219842	0	30.78555	32.38484	C2	C2	C2
Medan Tuntungan	23.75745	25.671	24.078	33.68152	C2	C1	C1
Medan Sunggal	40.20883	36.52396	8.046738	18.53225	C3	C3	C3
Medan Barat	22.31031	29.08608	3.671171	43.92544	C1	C1	C1
Medan Belawan	31.73983	26.28688	8.046738	15.63472	C3	C3	C3
Medan Marelan	36.60259	26.70206	18.10387	9.786615	C4	C4	C4
Medan Helvetia	38.2677	28.79236	16.72573	5.754221	C4	C4	C4
Medan Timur	50.9256	43.30127	18.72832	11.90705	C4	C4	C4

Table 10. Centroid cluster 1 iteration 3

	C1										
Subdistrict	a	b	c	d	e	f	G	h	I	j	k
Medan Polonia	0	55	0	56	50	9	25	0	0	0	0
Medan Johor	0	61	0	80	62	7	41	0	0	0	0
Medan Selayang	0	58	0	39	60	7	32	0	0	0	0
Medan Baru	0	67	0	55	64	7	51	0	0	0	0
Medan Tuntungan	0	76	0	76	80	8	29	0	0	0	0
Medan Barat	0	56	0	71	86	6	29	0	0	0	0

From the table above, it can be seen that the members of Cluster 1 are the districts of Meidan Polonia, Meidan Johor, Meidan Seilayang, Meidan Barui, Meidan Tuintuingan and Meidan Barat. And so on until centroid cluster 4, iteration 3 (Table 10).

Determines the maximum inter-cluster ratio

After finding all the ratios between the clusters, look for the maximum value of the ratio between the clusters.

Table 11. Maximum intercluisteir ratio

Ratio	1	2	3	4	Rmax
1	0	-1.54562	-0.35781	0.16362	0.16362
2	-1.54562	0	0.26138	0.447554	0.47554
3	-0.35781	0.26138	0	0.28152	0.28152
4	0.16362	0.847554	0.28152	0	0.28152

Calculate the Davies-Bouldin Index value.

Once we know the maximum value of the inter-cluster ratio, we can calculate the value of the Davieis-Bouildin Index.

$$DB = \frac{R1+R2+R3+R4}{4}$$

$$= \frac{0,16362+0,447554+0,28152+0,28152}{4}$$

$$= 0,30055$$

Based on the results and discussions described in the previous chapter, it can be seen that the city of Meidan has a group of areas prone to flood disasters. The area that is safe from flood disasters is Meidan Maimuin sub-district. Areas prone to flooding are the districts of Meidan Suinggal and Meidan Beilawan. The districts of Meidan Mareilan, Meidan Heilveitia and Meidan Timuir are prone to flooding. The areas most prone to flooding are the districts of Meidan Polonia, Meidan Johor, Meidan Seilayang, Meidan Barui, Meidan Tuintuingan, and Meidan Barat. From the calculations above, it can be concluded that the results of the clustering evaluation are suitable for using the Davieis-Bouildin Indeix, so the results obtained are around 0.30055. In this way, it can be concluded that the clues carried out can be stated as accurate.

Conclusion

Based on the results and discussions described in the previous chapter, it can be seen that the city of Meidan has a group of areas prone to flood disasters. The area that is safe from flood disasters is Meidan Maimuin sub-district. Areas prone to flooding are the districts of Meidan Suinggal and Meidan Beilawan. The districts of Meidan Mareilan, Meidan Heilveitia and Meidan Timuir are prone to flooding. The areas most prone to flooding are the districts of Meidan Polonia, Meidan Johor, Meidan Seilayang, Meidan Barui, Meidan Tuintuingan, and Meidan Barat. Based on the conclusions above, it is already clear that the Government should be more vigilant and pay more attention to the methods that can be used to mitigate flood disasters according to their ideals.

References

[1] Heryana, A. (2020). Pengertian dan Jenis-Jenis Bencana. *Jakarta: Universitas Esa Unggul*..

[2] Nuryanti, N., Tanesib, J. L., & Warsito, A. (2018). Pemetaan Daerah Rawan Banjir dengan Penginderaan Jauh dan Sistem Informasi Geografis di Kecamatan Kupang Timur Kabupaten Kupang Provinsi Nusa Tenggara Timur. *Jurnal Fisika: Fisika Sains Dan Aplikasinya*, *3*(1), 73-79.

[3] Binawan, B. P., & Junaedi, D. (2017). Aplikasi GIS Klasifikasi Tingkat Kerawanan Banjir Wilayah Kabupaten Bandung Menggunakan Metode Weighted Product. *Indonesia Journal on Computing (Indo-JC)*, *2*(1), 59-70.

[4] Statistik, B. P. (2022). *Badan Pusat Statistik*.

[5] Tampubolon, A. O., Siagian, A., & Ardiani, F. (2018). *Uji Daya Terima dan Kandungan Zat Gizi Dodol Dengan Penambahan Terung Ungu (Solanum Melongena) Dan Kulit Buah Pepaya (Carica Papaya L)* (Doctoral dissertation).

[6] Pramuji, S. E., & Putri, V. S. (2020). Meninjau Efektivitas Penegakan Hukum Penataan Ruang dalam Rangka Mewujudkan Tertib Tata Ruang. *Jurnal Pertanahan*, *10*(1)..

[7] Alamsyah, B., Nuraini, C., & Suwarno, B. (2022). STRATEGI MANAJEMEN MITIGASI BENCANA PESISIR PANTAI TIMUR SUMATERA UTARA. *PUBLISH BUKU UNPRI PRESS ISBN*, 1-81.

[8] Yana, M. S., Setiawan, L., Ulfa, E. M., & Rusyana, A. (2018). Penerapan Metode K-Means dalam Pengelompokan Wilayah Menurut Intensitas Kejadian Bencana Alam di Indonesia Tahun 2013-2018. *Journal of Data Analysis*, *1*(2), 93-102.

[9] Swindiarto, V. T. P., Sarno, R., & Novitasari, D. C. (2018, September). Integration of fuzzy C-means clustering and TOPSIS (FCM-TOPSIS) with silhouette analysis for multi criteria parameter data. In *2018 International Seminar on Application for Technology of Information and Communication* (pp. 463-468). IEEE.

[10] Suyanto, D. (2017). Data Mining untuk klasifikasi dan klasterisasi data. *Bandung: Informatika Bandung*.

[11] Murti, M. A. W. K. (2017). Penerapan metode K-means clustering untuk mengelompokan potensi produksi buah–buahan di provinsi daerah istimewa yogyakarta. *Universitas Sanata Dharma*.

[12] Yulian, E. (2018). Text mining dengan k-means clustering pada tema LGBT dalam arsip tweet masyarakat Kota Bandung. *Jurnal Matematika "MANTIK*, *4*(1), 53-58.

[13] Handoyo, R., Mangkudjaja, R., & Nasution, S. M. (2014). Perbandingan metode clustering menggunakan metode Single Linkage dan K-means pada Pengelompokan Dokumen. *Jurnal Sifo Mikroskil*, *15*(2), 73-82.

[15] Saikhu, A., & Gita, Y. B. (2013). Implementasi Deteksi Outlier pada Algoritma Hierarchical Clustering. *SEMNASTEKNOMEDIA ONLINE*, *1*(1), 07-45.

[16] Dehotman, B. J. (2018). *Peningkatan hasil evaluasi clustering davies-bouldin index dengan penentuan titik pusat cluster awal algoritma k-means* (Doctoral dissertation, Thesis, Universitas Sumatera Utara, Medan).."

[17] Renard, F. (2017). Flood risk management centred on clusters of territorial vulnerability. *Geomatics, Natural Hazards and Risk*, *8*(2), 525-543.

[18] Meylimdra Arini Permatadevi, Rully A. Hendrawan, S.Kom, M.Eng. 2013. "Kabel, Karakteristik Pelanggan Telepon Mojokerto), Menggunakan Clustering Som Dan K-Means Untuk Mengurangi Kesalahan Klasifikasi Pelanggan Perusahaan Telekomunikasi (Studi Kasus : Pt. Telkom." *Jurnal Teknik Pomits* 1(1).

[19] Sari, N. N. (2018). K-Affinity Propagation (K-AP) Clustering Untuk Klasifikasi Gempa Bumi (Studi Kasus: Gempa Bumi di Indonesia Tahun 2017).

[20] Hardaningrum, O., Sulaeman, C., & Supriyana, E. (2016). Zonasi Rawan Bencana Gempa Bumi Kota Malang Berdasarkan Analisis Horizontal Vertical To Spectral Ratio (HVSR)". In *dalam Prosiding Seminar Nasional MIPA* (pp. 170-175).