

K-Affinity Propagation Clustering and GIS on Instagram Accounts of Tourist Destinations in Java

Anugrah Sandra Aprilliana¹, Muhammad Muhajir^{2*}

¹ Department of Statistics, Islamic University of Indonesia
Sleman, Yogyakarta, Indonesia
E-mail: 15611145@students.uii.ac.id

² Department of Statistics, Islamic University of Indonesia
Sleman, Yogyakarta, Indonesia
E-mail: mmuhajir@uui.ac.id

Abstract

This research was conducted at the Department of Statistics of Indonesian Islamic Universities. The purpose of this study was to determine the general description of tourist destinations, the formation of tourist destination clusters, and the application of GIS. The tourist destination data used in this study were obtained from Instagram accounts on Java from January to December 2018. From each of these account posts there is a like value that researchers use as a reference for the most popular tourist destinations. The analytical methods used in this study are descriptive analysis, K-Affinity Propagation Clustering and GIS application. Obtained research results that in general tourist destinations are divided into 3 categories, namely nature, culture, and man-made. Grouping is formed using 3 clusters in each Instagram account to group each tourist destination from the level of the lowest number of likes to the highest. Then, with the application of GIS a visualization of mapping tourist destinations will be formed in Java in 2018. This research is expected to be able to provide benefits to the government especially the ministry of tourism in order to increase promotion through social media and review services related to tourist destinations offered on Java.

Keywords: Instagram, Tourist Destinations, Cluster, SIG, K-Affinity Propagation Clustering.

1. Introduction

Technology has an important role in life. One impact of technological development is the internet. One of the internet developments that is very loved in this era is social media. One social media that has many users is Instagram. Based on research conducted by We Are Social and Hootsuite, in the report "Digital 2019" Instagram occupies the 4th position as the most active social media platform used in Indonesia during the period January 2018 to January 2019. The era of social media like today, tourism activities is something that must be done because it can increase self-existence.

Everybody now competes to go to popular tourist attractions, show attendance with check-in facilities and capture interesting photos from every place visited, and of course all track record of tourist activities will be uploaded on social media [1].

Speaking of tourism, in Indonesia, especially Java offers a wide range of tourist destinations that can be visited. Which can be visited tourist destinations such as beaches, historic old buildings, churches, temples, museums, a wide variety of waterfalls, recreational parks, and so on.

1.1. Instagram Account

Instagram contains about a variety of account content, ranging from online shop accounts that sell goods and services, campus organization accounts and also Instagram accounts that promote tourism in a country or city [2]. Tourism is a travel activity that is carried out for a while from the original residence to the destination with the reason not to settle or make a living but only for fun, fulfill curiosity, spend leisure or holiday time and other destinations [3].

Table 1. Accounts Java Tourist Destinations

No	Accounts Instagram	Post	Followers	Following
1	Exploresemarang	2372	305K	69
2	Explorebandung	2504	711K	17
3	Exploremalang	2163	404K	7
4	Exploreserang	533	22,9K	66
5	Explore_jakarta	946	128K	27
6	Explorejogja	1928	655K	0

Source: <https://www.instagram.com/?hl=id> [4]

Table 1 is an Instagram account which contains photos about some interesting places in Java island. The number of tourist destinations found on the island of Java, makes researchers want to know the general description of tourist destinations based on Instagram accounts in table I. From each posting of these accounts there is a like value that researchers use as a reference for the most popular tourist destinations. Like in this case is assumed to be an impulse from the outside (existence) of Instagram users especially followers of the Instagram account.

1.2. Clustering

In addition, grouping of tourist destinations located on Java Island in 2018 is based on Instagram accounts *exploresemarang*, *explorebandung*, *exploremalang*, *exploreserang*, *explore_jakarta*, and *explorejogja*. The grouping method that will be used in this study is the K-Affinity Propagation Clustering method. The grouping in question uses data from the number of likes in the posts in each account. It aims to grouping each tourist destination from the level of the lowest number of likes to the highest.

1.3. Application of Geographic Information Systems

In the final step, visualization of mapping tourist destinations in Java Island in 2018 will be displayed based on Instagram accounts *exploresemarang*, *explorebandung*, *exploremalang*, *exploreserang*, *explore_jakarta*, and *explorejogja*. It aims to display a dynamic map connected to the internet so that it can bring up the geographical location of the tourist destination.

The rest of this paper is organized as follow: Section II describes the proposed method. Section III presents the obtained results and following by discussion. Finally, Section IV concludes this work.

2. Method

2.1. Descriptive Statistics

Descriptive statistics are methods relating to the collection and presentation of a group of data so as to provide useful information [5]. A descriptive statistic (in the count noun sense) is a summary statistic that quantitatively describes or summarizes features of a collection of information, while descriptive statistics (in the mass noun sense) is the process of using and analysing those statistics [6].

2.2. Geographic Information Systems (GIS)

Geographic Information System (GIS) is a system designed to capture, store, manipulate, analyze, organize and display all kinds of geographic data. SIG can not be separated from the spatial data, which is a reference to the position data, objects and their relationships within the earth. Spatial data is one of the items of information in which there is information on the earth including the surface of the earth, beneath the earth's surface, aquatic, marine and lower atmosphere [7]. The term geographic information system (GIS) means, in general, a software system designed to quantitatively represent, archive, manipulate, reason, visualize, and process spatial information [8].

2.3. K-Affinity Propagation Clustering (K-AP Clustering)

The K-Affinity Propagation Clustering method is an Affinity Propagation Clustering method that is modified to produce the optimal number of exemplar through Affinity Propagation Clustering. K-Affinity Propagation Clustering can produce K clusters as specified by the

user by adding constraints in the process of exchanging messages to limit the number of K clusters. K-Affinity Propagation Clustering while maintaining the quality of clustering from Affinity Propagation Clustering in terms of distortion. Another advantage of K-Affinity Propagation Clustering compared to Affinity Propagation Clustering is the trust in 1 item of data to be exemplar automatically to K-Affinity Propagation Clustering, while trust in Affinity Propagation Clustering is a user-defined parameter. The steps to start the K-Affinity Propagation Clustering algorithm are as follows [9]:

1. Similarities [10]:

$$s(i, k) = - \| x_i - x_k \|^2$$

2. Initialize availabilities matrix and confidence matrix:

Availabilities:

$$\forall i, k : a(i, k) = 0$$

Confidence:

$$\forall i : \eta^{out}(i) = \min(s)$$

3. Update responsibilities $\forall i, k :$

$$r(i, k) = s(i, k) - \max\{\eta^{out}(i) + a(i, i),$$

$$\max_{k': k' \neq i} \{s(i, k') + a(i, k')\}\}$$

$$r(i, i) = \eta^{out}(i) - \max_{k': k' \neq i} \{s(i, k') + a(i, k')\}$$

4. Update availabilities $\forall i, k :$

$$a(i, k) \leftarrow \min \{0, r(k, k) + \sum_{i': i' \notin \{i, k\}} \max \{0, r(i', k)\}\}$$

$$a(k, k) = \sum_{i': i' \neq k} \max \{0, r(i', k)\}$$

5. Update confidence, $\forall i :$

$$\eta^{in}(i) = a(i, i) - \max_{k': k' \neq i} \{s(i, k') + a(i, k')\}$$

$$\eta^{out}(i) = -R^K(\{\eta^{in}(k), k \neq i\})$$

Until converge

6. Output:

$$\text{Clustering assignments: } C = \{c_1, \dots, c_N\}$$

$$c_i = \arg \max_k \{a(i, k) + r(i, k)\}$$

3. Result

3.1. Descriptive Analysis of Top Tourist Destinations Java

Here is an overview of the tourism destination in Java Island based posts instagram account explorebandung, exploremalang, exploresemarang, explorejogja, explore_jakarta, and exploreserang:

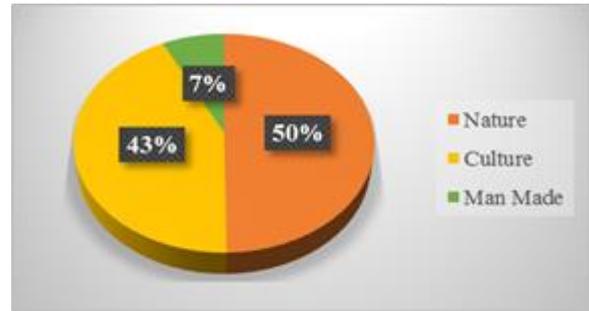


Fig. 1. Pie Chart The Tourist Destination Categories Of Java 2018

Fig. 1. a tourist destination on the island of Java in 2018. Of the total 410 existing tourist destinations, researchers divided them into three categories: nature, culture, and man-made. Distribution of tourist destinations category expressed in units of percent. Nature tourist destinations by category has a percentage of 50%, the category of culture has a percentage of 43%, and the category of man-made has a percentage of 7%. Category tourist destinations nature percentage much higher than the two other categories.

Discuss more details about the tourist destinations in the island of Java in 2018, based on the category of the tourist destinations in the fig. 1 researchers divided the tourist destination categories into sub categories. Sub division of this category is one attempt to pursuing the existing tourist destinations to be more specific. The result of the division of sub-categories of tourist destinations are as follows:

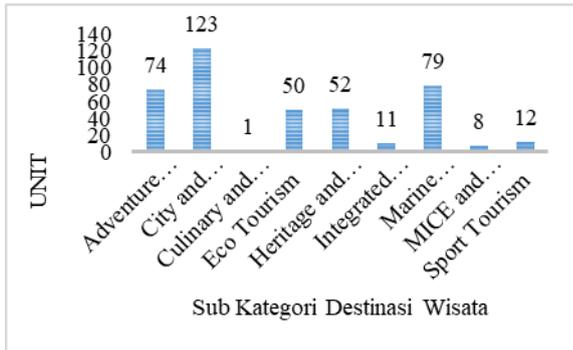


Fig. 2. Sub Category Bar Chart Java Tourist Destinations 2018

Based on Fig. 2. it can be seen that the number of tourist destinations sub category of adventure tourism as much as 74, city and village tourism as much as 123, culinary and shopping tourism as much as 1, eco-tourism as much as 50, heritage and pilgrim tourism as much as 52, the integrated area of tourism as much as 1, marine tourism as many as 79, MICE and events tourism as much as 8 and as much as 12 the sport tourism. Tourist Destinations with the highest number is the type of city and tourist destination tourism village of 123, while the tourist destinations with the lowest number is a kind of culinary tourist destinations and shopping tourism as much as 1.

Once you know the type of tourist destination in Java Island in 2018, following a bar chart of the 10 most popular tourist destinations on the island of Java in 2018 by instagram account explorebandung, exploremalang, exploresemarang, explorejogja, explore_jakarta, and exploreserang:

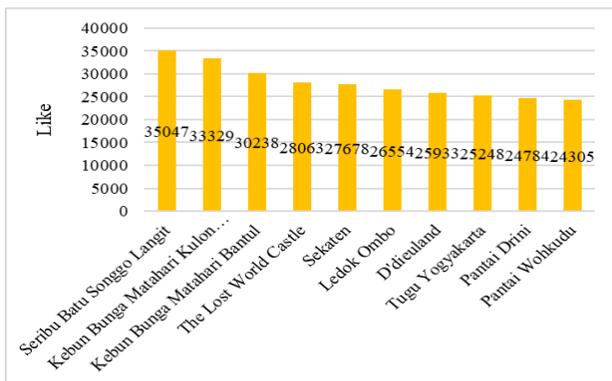


Fig. 3. Bar Chart Top Tourist Destinations in Java 2018

Fig. 3. is a bar chart graph that displays the 10 most popular tourist destinations in Java in 2018 with the highest number of likes based on Instagram accounts

explorebandung, exploremalang, exploresemarang, explore jogja, explore_jakarta, and exploreserang. The number 1 most popular tourist destination is Seribu Batu Songgo Langit which is a city and village tourism with 35047 likes, the second is Kulon Progo Sunflower Garden which is a city and village tourism with 33329 likes, the 3rd is Bantul Sunflower Garden which is a city and village tourism with 30238 likes, the 4th is The Lost World Castle which is a city and village tourism with 28063 likes, the 5th is Sekaten which is MICE and tourism events with 27678 likes, the 6th is Ledok Ombo which is adventure tourism with 26554 likes, the 7th is D'dieuland which is an integrated tourism area with 25933 likes, the 8th is Monument of Tugu Yogyakarta which is a city and village tourism with 25248 likes, the 9th is Drini Beach which is marine tourism with 24784 likes, and the 10th is Wokkudu Beach which is marine tourism with 24305 likes.

3.2. Analysis K- Affinity Propagation Clustering

K-Affinity Propagation Clustering in this study is used to group tourist destinations. The grouping is formed into 3 clusters to grouping tourist destinations on each instagram account explorebandung, exploremalang, exploresemarang, explorejogja, explore_jakarta, and exploreserang. In this study the grouping is done by using data from the number of likes on posts in each account. It aims to group each tourist destination from the level of the lowest number of likes to the highest.

The result of the grouping of the tourist destinations of each instagram account explorebandung, exploremalang, exploresemarang, explorejogja, explore_jakarta, and exploreserang using data on the number like the following:

Exploremalang

Based on the analysis using Software R, here are the results of grouping tourist destinations using K-Affinity Propagation Clustering algorithms:

Table 2. Results Exploremalang

Cluster	Eksemplar	Members of Tourist Destinations
1 (High)	Nirwana Hill	12
2 (Medium)	Savanna Beach	48
3 (Low)	Pasang Beach	59

Table 2. is the result of the analysis of K - Affinity Propagation Clustering. Grouping using the K-Affinity

Propagation Clustering method aims to group tourist destinations based on the number of likes from exploremalang Instagram accounts consisting of 119 tourist destinations. There are three clusters with each member and the different exemplars. Cluster 1 consists of 12 tourist destinations. As for tourist destinations which become exemplars cluster 1 is Nirvana Hill. Cluster 2 consists of 48 tourist destinations. As a tourist destination the exemplars of the cluster 2 is Savanna Beach. Cluster 3 consisted of 59 tourist destinations. As a tourist destination the exemplars of the cluster 3 is Pasang Beach.

Explore_jakarta

Based on the analysis using Software R, here are the results of grouping tourist destinations using K-Affinity Propagation Clustering algorithms:

Table 3. Results Explorejakarta

<i>Cluster</i>	<i>Eksemplar</i>	<i>Members of Tourist Destinations</i>
1 (High)	National Monument	6
2 (Medium)	Kebon Melati Reservoir	33
3 (Low)	Vihara Dharma Bhakti	18

Table 3. is the result of the analysis of K - Affinity Propagation Clustering. Grouping using the K - Affinity Propagation Clustering method aims to group tourist destinations based on the number of likes from the explore_jakarta Instagram account consisting of 57 tourist destinations. There are three clusters with each member and the different exemplars. Cluster 1 consisted of 6 tourist destinations. As for tourist destinations which become exemplars cluster 1 is the National Monument. Cluster 2 consists of 33 tourist destinations. As a tourist destination the exemplars of the cluster 2 is Kebon Melati Reservoir. Cluster 3 is comprised of 18 tourist destinations. As a tourist destination the exemplars of the cluster 3 is the Vihara Dharma Bhakti.

Exploresamarang

Based on the analysis using Software R, here are the results of grouping tourist destinations using K-Affinity Propagation Clustering algorithms:

Table 4. Results Exploresamarang

<i>Cluster</i>	<i>Eksemplar</i>	<i>Members of Tourist Destinations</i>
1 (High)	Ayana Gedong Songo	3
2 (Medium)	Benteng Fort Willem I	27
3 (Low)	Hill of Love Rawa Pening	40

Table 4. is the result of the analysis of K - Affinity Propagation Clustering. Grouping using the K - Affinity Propagation Clustering method aims to group tourist destinations based on the number of likes from the Instagram account exploresamarang which consists of 70 tourist destinations. There are three clusters with each member and the different exemplars. Cluster 1 consists of 3 tourist destinations. As for tourist destinations which become exemplars cluster 1 is Ayana Gedong Songo. Cluster 2 has 27 tourist destinations. As a tourist destination the exemplars of the cluster 2 is Fort Fort William I. Cluster 3 consisted of 40 tourist destinations. As a tourist destination the exemplars of the cluster 3 is the Hill of Love Rawa Pening.

Exploreserang

Based on the analysis using Software R, here are the results of grouping tourist destinations using K-Affinity Propagation Clustering algorithms:

Table 5. Results Exploreserang

<i>Cluster</i>	<i>Eksemplar</i>	<i>Members of Tourist Destinations</i>
1 (High)	Mountain Pilar	10
2 (Medium)	Cibaja Mountain Nature Tourism	16
3 (Low)	Monument of the Struggle of the People of Banten	17

Table 5. is the result of the analysis of K - Affinity Propagation Clustering. Grouping using the K-Affinity Propagation Clustering method aims to group tourist destinations based on the number of likes from the Instagram account exploreserang which consist of 43 tourist destinations. There are three clusters with each member and the different exemplars. Cluster 1 consists of 10 tourist destinations. As for tourist destinations which become exemplars cluster 1 is Mountain Pillar. Cluster 2 consists of 16 tourist destinations. As a tourist destination the exemplars of the cluster 2 is Cibaja Mountain Nature Tourism. Cluster 3 consists of 17 tourist destinations. As a tourist destination the

exemplars of the cluster 3 is the Monument of the Struggle of the People of Banten.

Explorebandung

Based on the analysis using Software R, here are the results of grouping tourist destinations using K-Affinity Propagation Clustering algorithms:

Table 6. Results Explorebandung

<i>Cluster</i>	<i>Eksemplar</i>	<i>Members of Tourist Destinations</i>
1 (High)	Sanghyang Heuleut	8
2 (Medium)	West Java Struggle Monument	18
3 (Low)	Regol Square	13

Table 6. is the result of the analysis of K - Affinity Propagation Clustering. Grouping using the K-Affinity Propagation Clustering method aims to group tourist destinations based on the number of likes from the Instagram account explorebandung consisting of 39 tourist destinations. There are three clusters with each member and the different exemplars. Cluster 1 comprised 8 tourist destinations. As for tourist destinations which become exemplars cluster 1 is Sanghyang Heuleut. Cluster 2 is comprised of 18 tourist destinations. As a tourist destination the exemplars of the cluster 2 is the West Java Struggle Monument. Cluster 3 consisted of 13 tourist destinations. As a tourist destination the exemplars of the cluster 3 is Regol Square.

Explorejogja

Based on the analysis using Software R, here are the results of grouping tourist destinations using K-Affinity Propagation Clustering algorithms:

Table 7. Results Explorejogja

<i>Cluster</i>	<i>Eksemplar</i>	<i>Members of Tourist Destinations</i>
1 (High)	Bantul Sunflower Garden	5
2 (Medium)	Cemara Sewu Beach	37
3 (Low)	Wediombo Beach	40

Table 7. is the result of the analysis of K - Affinity Propagation Clustering. Grouping using the K-Affinity Propagation Clustering method aims to group tourist destinations based on the number of likes from the Instagram account explorejogja which consists of 82

tourist destinations. There are three clusters with each member and the different exemplars. Cluster 1 consists of five tourist destinations. As for tourist destinations which become exemplars cluster 1 are the Bantul Sunflower Garden, cluster 2 consists of 37 tourist destinations. As a tourist destination the exemplars of cluster 2 is a Cemara Sewu Beach. Cluster 3 consisted of 40 tourist destinations. As a tourist destination the exemplars of the cluster 3 is Wediombo Beach.

Validation Cluster

In this study the cluster validation process is done by comparing the cluster variance values of K-AP and K-Medoids. A good method is a method with a small cluster variance.

Table 8. Clustering Result K-AP and K-Medoids

<i>Instagram Explore Account</i>	<i>Method</i>	<i>Cluster Variance</i>
exploremaalang	K- AP	0.421
	K- Medoids	10.422
explore_jakarta	K- AP	0.265
	K- Medoids	0.296
exploresemarang	K- AP	0.226
	K- Medoids	0.447
exploreserang	K- AP	0.390
	K- Medoids	0.390
explorebandung	K- AP	0.309
	K- Medoids	0.314
explorejogja	K- AP	0.308
	K- Medoids	0.439

Based on the validation of K-AP and K-Medoids, it can be seen that the K-AP method is the best method when viewed from cluster variance.

3.3. Application of Geographic Information Systems

The application of geographic information systems is used to display visualization of tourist destinations in the form of maps. In its application, the geographical information system in this study aims to display maps connected to the internet so that it can bring up the geographical location of tourist destinations based on Instagram accounts exploremaalang, explore_jakarta, exploreserang, exploresemarang, explorebandung, and explorejogja. The visualization of the mapping of tourist destinations are:

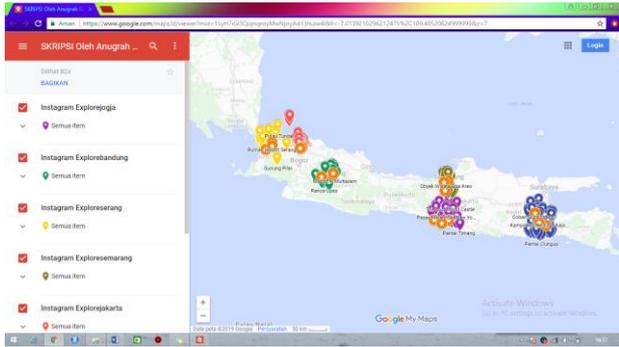


Fig. 4. Visualization Distribution Tourist Destinations

Fig. 4. is a visualization of the distribution of tourist destinations in Java based on posts on instagram accounts exploremalang, explorers_jakarta, exploresemarang, exploreserang, explorebandung, and explorejogja in 2018. Blue coordinate points are tourist destinations in Malang based on exploremalang instagram accounts. The brown coordinate point is a tourist destination in Semarang based on the exploresemarang account posting. The purple coordinate point is a tourist destination in D.I Yogyakarta based on Explorejogja Instagram account posts. The green coordinate point is a tourist destination in Bandung based on explorebandung Instagram account posts. The pink coordinate point is a tourist destination in DKI Jakarta based on the explore_jakarta Instagram account post. The yellow coordinate is a tourist destination located in Serang based on Instagram exploreserang account posts.

In addition to the colors mentioned above, there are coordinate points with colors and icons that are different from others. The coordinates are orange and use the star as the icon. The coordinates are used to indicate 3 tourist destinations with the highest number of likes in each region. To see a map and geographical information of tourist destinations as shown in fig. 4. you can access the following link:

http://bit.ly/MapsofTouristDestinationin_Java

4. Conclusion

There are as many as 410 tourist destinations in Java in 2018 by posting instagram account explorebandung, exploremalang, exploresemarang, explorejogja, explore_jakarta, and exploreserang. In general, tourist destinations are divided into 3 categories namely nature, culture, and man-made. Nature tourist destinations by category has a percentage of 50%, the category of culture has a percentage of 43%, and the category of man-made has a percentage of 7%.

The K - Affinity Propagation Clustering method is used to grouping tourist destinations on Java Island in 2018. Three clusters are used to grouping tourist destinations on each Instagram account explorebandung, exploremalang, exploresemarang, explorejogja, explore_jakarta, and exploreserang. The groupings referred to in this study are groupings made using data from the number of likes on posts in each account. It aims to group each tourist destination from the level of the lowest number of likes to the highest. Grouping is done separately for each Instagram account, so that the results obtained cluster from each Instagram account explorebandung, exploremalang, exploresemarang, explore jogja, explore_jakarta, and exploreserang. Visualization of mapping tourist destinations in Java in 2018 based on Instagram account explorebandung, exploremalang, exploresemarang, explorejogja, explore_jakarta, and exploreserang accounts can be accessed through the following link:

http://bit.ly/MapsofTouristDestinationin_Java

References

- [1] Iswandhani N., Visualisasi Pemetaan Destinasi Wisata Provinsi Yogyakarta dengan Metode *Cluster K-means* dan Penerapan Teori Graf Pada Analisis Spasial, (Yogyakarta: Respository Universitas Islam Indonesia), 2017.
- [2] Yotenka, R., Muhajir, M., Iswandhani, N., and Ima, H., Tourist Destination Mapping Visualization in Yogyakarta using Affinity Propagation and Social Network Analysis: A Case Study on @explorejogja instagram in 2017, *Advances in Social Science, Education and Humanities Research (ICEBESS)*, 2018, volume 313.
- [3] Suwena, I.K. dan Widyatmaja, I., G., N., *Pengetahuan Dasar Ilmu Pariwisata*, Denpasar: Pustaka Larasan, 2017.
- [4] Instagram, (2010, October 6), Accessed on January, 29 2018, from
- [5] <https://www.instagram.com/?hl=id>
- [6] Walpole, Ronald E., *Introductions to Statistics*, Jakarta: Gramedia Pustaka Utama, 1995.
- [7] Mann, Prem S., *Introductory Statistics 2nd ed.*, Wiley ISBN 0-471-31009-3, 1995.

- [8] Irwansyah E., Sistem Informasi Geografis: Prinsip Dasar dan Pengembangan Aplikasi, Yogyakarta: Digibooks, 2013.
- [9] Longley, P., A., Goodchild, M., F., Maguire, D., J., and Rhind, D., W., Geographic Information Systems and Science 3rd ed., John Wiley & Sons, Inc., ISBN 978-0-470-72144-5, 2011.
- [10] Zhang X., W. Wei, N. Kjetil, and S. Michele, K-AP: Generating Specified K Clusters by Efficient Affinity Propagation, Data Mining (ICDM), 2010 IEEE 10th International Conference on 2010.
- [11] Frey, B., J., and Dueck, D, Clustering by Passing Messages Between Data Points, Journal of Sciencemag, volume 315, 2007