

IMPLEMENTASI K-MEANS CLUSTERING PADA APLIKASI GIS (GEOGRAPHIC INFORMATION SYSTEM) (STUDI KASUS PERTANIAN PADI)

Jasman Pardede¹, Milda Gustiana², Muadz Nurhasan³

- (1) Institut Teknologi Nasional Bandung, jasman@itenas.ac.id
(2) Institut Teknologi Nasional Bandung, mghusada@yahoo.com
(3) Institut Teknologi Nasional Bandung, zenkmuadz@gmail.com

Abstrak

K-Means merupakan pemodelan pengelompokan data sesuai dengan karakteristik atau ciri-ciri bersama yang serupa. Pengelompokan data yang digunakan dalam penelitian ini diterapkan pada penentuan hasil dari pertanian dalam setiap daerah dengan mengelompokkan daerah sukses panen, kurang tercapai dan gagal panen, dengan parameter yang ditentukan yaitu persentase luas panen dan persentase produksi dari luas tanam (ha), luas panen (ha) dan produksi (ton). Pengelompokan hasil pertanian bertujuan untuk menginformasikan kepada masyarakat dan pemerintah daerah – daerah yang berpotensi untuk pertanian padi. Dalam penelitian ini daerah pengelompokan yang sudah didapat dimunculkan ke dalam map untuk bisa menyajikan suatu Geographic Information System (GIS) untuk pertanian padi. Pengimplementasian metode K-Means terletak pada pencocokan centroid dan perhitungan jarak untuk mendapatkan jarak terdekat dengan menggunakan Euclidean Distance. Berdasarkan pengujian dari 16 data tahun 2012 - 2015 yang telah dilakukan, tingkat keakuratan pengelompokan suatu cluster dengan menggunakan metode K-Means mencapai 78.125%.

Kata kunci: K-Means, Geographic Information System

Abstract

K-Means clustering is modeling data in accordance with the characteristics or traits shared similar. Grouping the data used in this study applied to the determination of the result of agriculture in each region by grouping the area successful harvest , less accomplished and failed harvests , with the specified parameters , namely percentage of area harvested and the percentage of production from the planting area (ha) , the harvested area (ha) and production (ton). Grouping of harvest aims to uniform peoples and government the areas that have potential for farming. In this study, regional grouping which has been obtained will be raised to the folder to be able to present a Geographic Information System (GIS) for agriculture . Implementation of the K -Means method lies in matching centroid and distance calculations to get the closest distance by using Euclidean Distance . Based on testing of 16 data that has been done , the accuracy of grouping a cluster using the K -Means reached 78.125 % .

Keywords: K-Means clustering, GIS, Euclidean Distance.

1. PENDAHULUAN

1.1 Latar Belakang

GIS (Geographic Information System) adalah suatu sistem informasi yang mengelola data yang memiliki informasi spasial yang memiliki kemampuan untuk membangun, menyimpan, mengelola dan menampilkan informasi berefrensi geografis[3], seperti halnya google map yang menampilkan sebuah informasi geografis dari citra satelit untuk mengetahui peta permukaan bumi, hanya saja pada map yang di sediakan oleh google hanya menyediakan informasi yang menampilkan sebuah potret peta dunia yang hanya bisa melihat suatu nama jalan dan

wilayah/daerah.

Bagaimana jika membuat suatu peta yang menampilkan daerah pertanian dalam suatu kota sehingga bisa mengetahui hasil pertanian yang sudah dilakukan dari beberapa daerah, maka dari itu agar lebih mudah mengetahui hasil pertanian dibutuhkan suatu pengelompokan hasil pertanian, dan untuk menentukan suatu pengelompokan tersebut dibutuhkan suatu metode pengelompokan hasil pertanian yaitu dengan metode K-Means Clustering. K-Means Clustering merupakan salah satu metode data clustering non hirarki

yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih kelompok (cluster)[6].

Sehingga dari hasil proses *K-means* ini mendapatkan sebuah pengelompokan pertanian yang diperoleh untuk dipetakan ke dalam GIS, dengan begitu hal ini bisa menginformasikan kepada user khususnya masyarakat dan pemerintah guna mengetahui potensial daerah mana saja yang cocok untuk penanaman padi dalam daerah tersebut[2] ataupun untuk mengatasi dalam berbagai kendala dalam produktivitas setelah melihat map dari pertanian.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana pengimplementasian proses *K-Means Clustering* ke dalam sistem.
2. Bagaimana keakuratan pengelompokan hasil *cluster* dari *K-Means Clustering*.
3. Bagaimana sistem dapat memetakan peta daerah pertanian padi.

1.3 Tujuan

Tujuan dari penelitian ini adalah membuat *Geographic Information System* hasil pertanian padi dengan mengimplementasikan *K-Means*, dimana sistem ini juga menguji ketepatan akurasi informasi pengelompokan data hasil pertanian padi dengan menggunakan metode *K-Means Clustering*.

1.4 Metode Penelitian

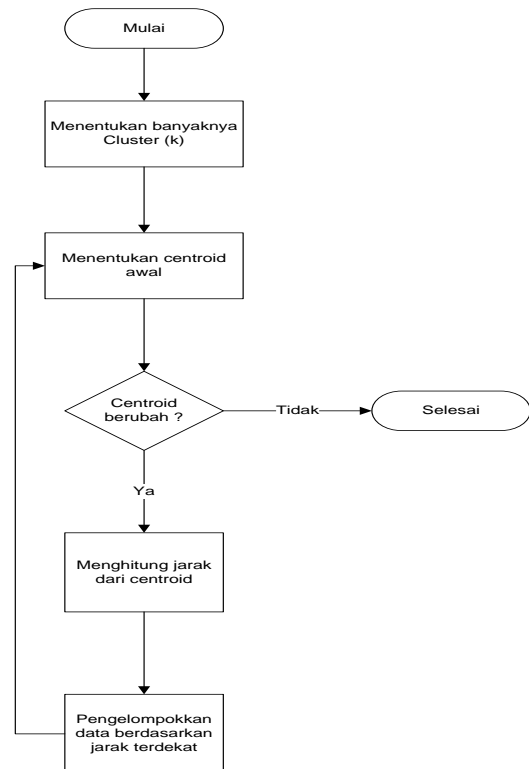
Pada penelitian ini dilakukan dengan menggunakan tahapan penelitian sebagai berikut:

1. Metode Pengumpulan Data
Pengumpulan data dilakukan dengan cara sebagai berikut :
 - a. Studi literatur, yaitu dengan mengumpulkan data dan informasi mengenai permasalahan yang dibahas dan mempelajari literatur-literatur berupa buku, jurnal, artikel ilmiah, serta website.
 - b. Survey, yaitu dengan mengumpulkan data yang dibutuhkan untuk daerah - daerah pertanian dan hasil produksi didapatkan dengan survey langsung dari pusat data Dinas Pertanian dan Ketahanan Pangan di Kota Bandung yang kemudian data akan diproses ke dalam sistem.
2. Metode Pengembangan Sistem
Metode pengembangan sistem yang digunakan dalam tugas akhir ini adalah model *Prototype*. Metode ini dipilih karena sistem yang akan dibuat berskala kecil dalam pengembangan sistemnya sehingga dalam pembuatannya akan terselesaikan dengan cepat. Metode ini juga hanya menganalisa bagian – bagian yang penting saja dengan membuat fungsi yang sederhana atau tidak besar sehingga bisa menghemat jangka waktu dalam pembuatan sistem.

2. Metodologi

2.1 K-Means^[1]

K-means merupakan salah satu metode *clustering* non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*. Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam *cluster* yang lain. Algoritma untuk melakukan *K-Means clustering* digambarkan pada Gambar 1.



Gambar 1. Algoritma *K-means*

Keterangan :

1. Menentukan banyaknya *Cluster* (k).
2. Pilih titik *centroid* secara acak.
3. Menghitung jarak dengan *Euclidean Distance*. *Euclidean distance* adalah perhitungan untuk menentukan jarak dari 2 buah titik. Rumus *Euclidean distance* bisa dihitung menggunakan persamaan (1).

$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2} \quad \dots\dots(1)$$

Dimana:

d_{ij} = Jarak objek antara objek i dan j

P = Dimensi data

x_{ik} = Koordinat dari obyek i pada dimensi k

x_{jk} = Koordinat dari obyek j pada dimensi k

4. Kelompokkan data sehingga terbentuk K buah *cluster* dengan titik *centroid* dari setiap *cluster* merupakan titik *centroid* yang telah dipilih sebelumnya

5. Perbaharui nilai titik *centroid*
6. Ulangi langkah 2 dan 3 sampai nilai dari titik *centroid* tidak lagi berubah.

3. Pembahasan

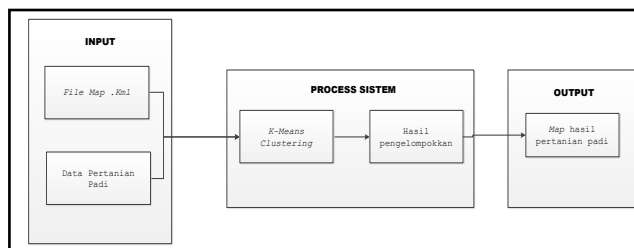
3.1 Analisis kebutuhan sistem

Hasil analisis kebutuhan sistem untuk diproses agar sistem berjalan sesuai dengan yang diharapkan dengan pengumpulan data pertanian yang diperoleh dari wawancara dan survey ke Dinas Pertanian Dan Ketahanan Pangan Kota Bandung didapatkan data hasil pertanian dan beberapa kesimpulan yaitu :

1. Data hasil pertanian yang diperoleh dari tahun 2012 sampai tahun 2015 yaitu luas tanam (ha), luas panen (ha) dan produksi (ton).
2. Penentuan parameter data yang diproses adalah persentase luas panen dan persentase produksi yang didapat dari data pertanian luas panen (ha), luas panen (ha), dan produksi (ton).
3. Dalam penentuan *cluster* dibagi menjadi 3 pengelompokkan hasil pertanian yaitu :
 - a. *Cluster* 1 (Sukses Panen).
 - b. *Cluster* 2 (Kurang Tercapai).
 - c. *Cluster* 3 (Gagal Panen).
4. Dalam penentuan titik *centroid* awal ditentukan dari data persentase luas panen dan persentase produksi berdasarkan *range* sebagai berikut :
 - a. Titik *centroid Cluster* 1 (*Range* 100 % – 70 %).
 - b. Titik *centroid Cluster* 2 (*Range* 11% – 71 %).
 - c. Titik *centroid Cluster* 3 (*Range* 0 % – 10 %).

3.2 Blok Diagram

Blok diagram dari proses sebuah sistem yang dibuat ditampilkan pada Gambar 2, dimana dalam *process input* dibutuhkan suatu data yang terdiri dari *file map* yang sudah terdigitasi serta sudah dilakukan *editing map*, *editing* data attribut dan dibutuhkan data pertanian untuk kemudian diproses dalam *process* sistem dengan menggunakan *k-means clustering* untuk mengelompokkan data menjadi daerah pengelompokkan hasil pertanian yang kemudian hasil pengelompokkan dipetakan ke dalam *map* dalam bentuk *web gis*.

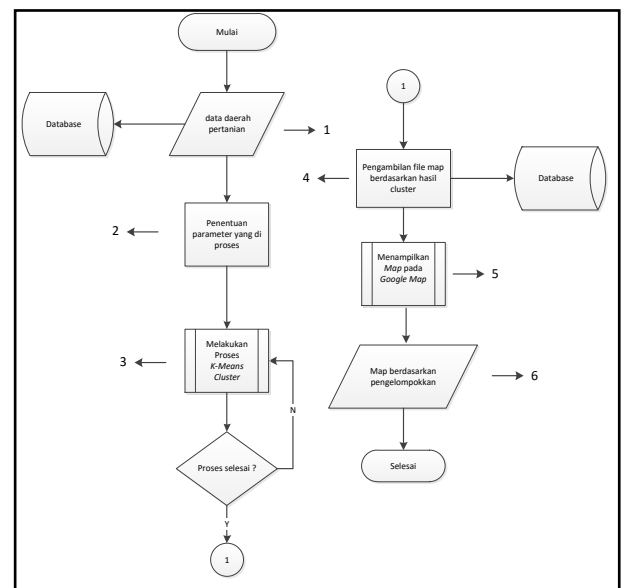


Gambar 2. Blok Diagram

3.3 Alur Sistem

Alur proses sistem yang dibuat ditampilkan pada Gambar 3 dalam bentuk *flowchart*, dimana :

1. Inisiasi data daerah pertanian yang diambil dari *database* meliputi luas tanam (ha), luas panen (ha) dan produksi (ton).
2. Setelah data diambil, menentukan parameter yang diproses dalam *k-means* berdasarkan data - data yang diperoleh yaitu persentase dari luas panen (%) dan persentase dari produksi (%).
3. Setelah menentukan parameter dari data yang sudah diperoleh selanjutnya melakukan proses *cluster* pengelompokkan dengan metode *K-Means Clustering*.
4. Setelah proses *k-means clustering* selesai, maka didapatlah hasil kategori pengelompokkan daerah pertanian yang ditentukan, kemudian pengambilan *map file* berdasarkan hasil proses *k-means* dari *database*.
5. Kemudian *map file* yang sudah didapat akan dimunculkan dan dipetakan ke dalam *google map*.
6. Keluaran yang dihasilkan adalah daerah - daerah hasil pengelompokkan yang sudah dipetakan ke dalam *map virtual* (*Google map*).



Gambar 3. Alur sistem

3.4 Alur Implementasi K-means

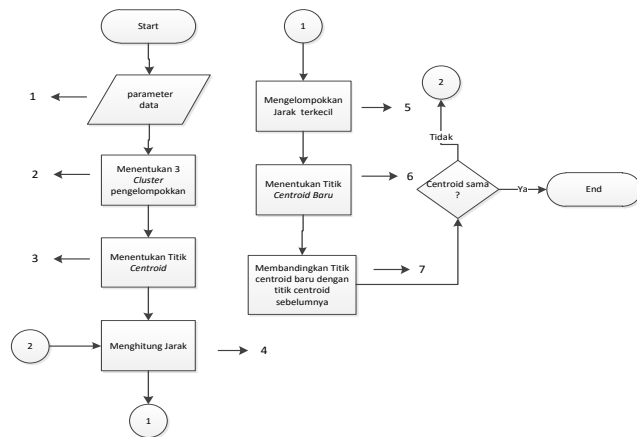
Proses implementasi *K-Means* dari sebuah sistem yang dibuat digambarkan oleh Gambar 4. Proses alur dari implemetasi *K-Means* digambarkan pada Gambar 5, sebagai berikut :

1. Inisialisasi parameter data diproses yaitu persentase dari luas panen dan persentase dari produksi.

$$\text{Persentase luas panen} = \frac{\text{Luas Panen}}{\text{Luas Tanam}} \times 100$$

$$\text{Persentase produksi} = \frac{\text{produktivitas}}{\text{Luas Tanam}} \times 100$$

$$\text{Ket : Produktivitas} = 6,5 \text{ ton}$$



Gambar 4. Implementasi K-Means

Tabel 1. Data 2012

Daerah	Luas Tanam (ha)	Luas Panen (ha)	Produksi (ton)	Persentase Luas Panen	Persentase Produksi
Antapani	57	16	100	28.070175	26.9905533
Arcamani k	94	84	546	89.361702	89.3617021
Babakan Ciparay	42	39	253	92.857142	92.6739926
Bandung Kidul	158	133	864	84.177215	84.1285297
Bandung Kulon	27	8	52	29.629629	29.6296296
Bojongloa Kidul	15	0	0	0	0
Buahbatu	171	119	773	69.590643	69.5456590
Cibiru	160	157	1020	98.125	98.0769230
Cinambo	239	156	1027	65.271966	66.1087866
Gedebage	900	829	5388	92.111111	92.1025641
Kiaradong	136	58	377	42.647058	42.6470588
Mandalajati	33	18	117	54.545454	54.5454545
Panyileukan	128	75	487	58.59375	58.5336538
Rancasari	261	133	864	50.957854	50.9283819
Regol	7	4	26	57.142857	57.1428571
Ujungberung	380	248	1612	65.263157	65.2631578

- Menentukan 3 cluster pengelompokan daerah pertanian yaitu sukses panen, kurang tercapai dan gagal panen, untuk range digunakan dalam penentuan titik centroid awal bisa dilihat pada Tabel 2.

Tabel 2. Cluster

Cluster	Keterangan	Range
Cluster 1	Sukses Panen	71% - 100%
Cluster 2	Kurang Tercapai	11% - 70%
Cluster 3	Gagal Panen	0% - 10%

Ket : untuk penentuan range didapat dari analisis kesimpulan pengumpulan data.

- Menentukan titik centroid awal dengan data dari persentase luas panen. Penentuan titik centroid awal yang dilakukan adalah dengan mengambil nilai acak berdasarkan range dari persentase luas panen, untuk variable x adalah titik centroid dari parameter persentase luas panen dan variable y adalah titik centroid dari parameter persentase produksi yang di perlihatkan pada Tabel 3.

Tabel 3. Centroid awal

Centroid Awal			
Keterangan	X	Y	Daerah
Cluster 1	98.125	98.0769230	Cibiru
Cluster 2	29.6296	29.629629	Bandung Kulon
Cluster 3	0	0	Bojongloa Kidul

- Menghitung jarak dari tiap cluster dengan menggunakan Euclidean distance. Menghitung jarak dengan rumus Euclidean Distance untuk menentukan jarak dari masing – masing cluster yang sudah ditentukan dengan menggunakan persamaan (1), seperti pada Tabel 4.
- Mengelompokkan jarak yang terkecil dari ke 3 jarak dari cluster yang paling minimum dari suatu cluster yang diperlihatkan pada Tabel 5.

Tabel 4. Penentuan jarak

Daerah	Jarak C1	Jarak C2	Jarak C3
Antapani	99.80456107	3.06539087459	38.941298346821.
Arcamanik	12.35922593	84.4739070353	126.37653110568.
Babakan Ciparay	7.545990711	89.2877939060	131.19038797767.
Bandung Kidul	19.72557677	77.1075171376	119.01013850212.
Bandung Kulon	96.83309218668	5.2755106611775E-13.	41.902624070314.
Bojongloa Kidul	138.7357144	41.9026240703	0.
Buahbatu	40.351487487343.	56.481607747958.	98.38422800349.
Cibiru	8.5265128291212E-14.	96.83309218668.	138.73571445463.
Cinambo	45.839759570451.	51.001030122756.	92.902105904496.
Gedebage	8.4770174426153.	88.356115067267.	130.25873907109.
Kiaradong	78.423668674257.	18.409424913244.	60.312048983558.
Mandalajati	61.596798086678.	35.236297513673.	77.138921583987.
Panyileukan	55.914129413053.	40.918979517318.	82.82159242398.
Rancasari	66.6912	30.14183178025	72.04445166029.

	6294093 8.	8.	
Regol	57.9235 1670140 3.	38.90957949386 2.	80.81220356417 7.
Ujungberung	46.4396 7971447 4.	50.39341894771 9.	92.29604301803 4.

Tabel 5. Penentuan jarak minimum

Jarak C1	Jarak C2	Jarak C3	Suk ses	Kur ang	Ga gal
99.80456107 5578.	3.065390874 5978.	38.9412983 46821.		K	
12.35922593 8296.	84.47390703 5366.	126.376531 10568.	S		
7.545990711 395.	89.28779390 6025.	131.190387 97767.	S		
19.72557677 1253.	77.10751713 7635.	119.010138 50212.	S		
96.83309218 668.	5.275510661 1775E-13.	41.9026240 70314.		K	
138.7357144 5463.	41.90262407 0314.	0.			G
40.35148748 7343.	56.48160774 7958.	98.3842280 0349.	S		
8.526512829 1212E-14.	96.83309218 668.	138.735714 45463.	S		
45.83975957 0451.	51.00103012 2756.	92.9021059 04496.	S		
8.477017442 6153.	88.35611506 7267.	130.258739 07109.	S		
78.42366867 4257.	18.40942491 3244.	60.3120489 83558.		K	
61.59679808 6678.	35.23629751 3673.	77.1389215 83987.		K	
55.91412941 3053.	40.91897951 7318.	82.8215924 2398.		K	
66.69126294 0938.	30.14183178 0258.	72.0444516 6029.		K	
57.92351670 1403.	38.90957949 3862.	80.8122035 64177.		K	
46.43967971 4474.	50.39341894 7719.	92.2960430 18034.	S		

6. Menentukan titik *centroid* baru dengan menghitung rata – rata dari *cluster* yang sama dengan menggunakan persamaan (2) ditampilkan pada Tabel 6, variable x dan y pada persamaan (2) adalah nilai persentase luas panen nilai persentase produksi.

$$Centroid_i = \frac{(x_{i1} + x_{i2} + \dots + x_{in})}{n}, \frac{(y_{i1} + y_{i2} + \dots + y_{in})}{n} \dots (2)$$

7. Membandingkan titik *centroid* yang baru dengan titik *centroid* yang sebelumnya, jika titik *centroid* baru tidak sama dengan titik *centroid* yang sebelumnya maka akan ditentukan titik *centroid* baru, sampai titik *centroid* tidak berubah. Hasilnya seperti yang dinyatakan pada Tabel 7.

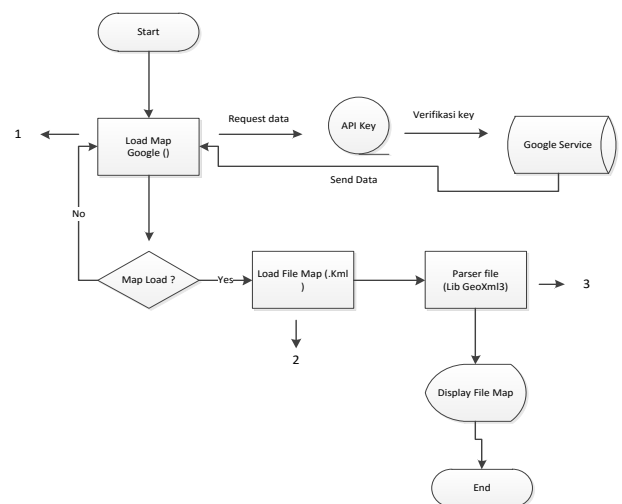
Tabel 6. Centroid baru

Keterangan	X	Y
Cluster 1	82.09474	82.15766
Cluster 2	45.9409	45.7739
Cluster 3	0	0

Tabel 7. Hasil akhir

Daerah	Cluster
Antapani	Kurang Tercapai
Arcamanik	Sukses Panen
Babakan Ciparay	Sukses Panen
Bandung Kidul	Sukses Panen
Bandung Kulon	Kurang Tercapai
Bojongloa Kidul	Gagal Panen
Buahbatu	Sukses Panen
Cibiru	Sukses Panen
Cinambo	Sukses Panen
Gedebage	Sukses Panen
Kiaracondong	Kurang Tercapai
Mandalajati	Kurang Tercapai
Panyileukan	Kurang Tercapai
Rancasari	Kurang Tercapai
Regol	Kurang Tercapai
Ujungberung	Sukses Panen

3.5 Proses menampilkan Map

**Gambar 5.** Alur menampilkan map

Proses menampilkan sebuah *map* yang sudah dilakukan digitasi ke dalam *google map* digambarkan oleh Gambar 5, sebagai berikut :

1. Untuk memunculkan *file map* yang sudah dibuat dengan digitasi ke dalam *google map*, pertama- tama *google map* harus terbuka terlebih dahulu.
2. Untuk membuka *google map* dibutuhkan suatu *API key*, *API key* adalah kode yang sudah disahkan oleh program komputer yang memanggil *API* (*Application Programming Interface*), dalam penelitian ini hanya menggunakan *API key* untuk mendapatkan data *Google Map* dari *google service* untuk dijadikan peta dasar.
3. Setelah sukses melakukan verifikasi pada *Google service* dan *google map* sudah terbuka, *file map* yang sudah terdigitasi dimunculkan ke dalam *google map* dengan bantuan *library geoxml3.js* untuk mengubah *file* yang berformat *.kml* menjadi peta kedalam *google map*.

3.6 Pengujian akurasi

Berdasarkan dari penelitian yang sudah dilakukan hasil penelitian dengan 16 daerah pengelompokkan dengan metode *k-means clustering* dari daerah 2012 – 2015 sudah didapatkan dan untuk mengetahui akurasi dari proses *K-Means* yang telah diujikan dari sistem yang sudah dibuat dengan membandingkan dari ke 16 data daerah yang sudah diujicobakan dengan proses *K-Means* dan dicocokkan dengan data asli yang ditentukan manual dapat dilihat dari persamaan (3), pada Tabel 8 untuk hasil akurasi dari tahun 2012-2015.

$$\text{Akurasi} = \frac{\text{data cocok}}{\text{jumlah data}} \times 100 \% \quad \dots\dots\dots (3)$$

Tabel 8. Hasil Akurasi

Tahun	Data cocok	Data tidak cocok	Akurasi
2012	13	3	81.25%
2013	12	4	75%
2014	13	3	81.25%
2015	12	4	75%

Maka, akurasi dari ke 16 data dari tahun 2012 – 2015 yang diujicobakan dan dirata-ratakan didapat hasil akurasi sebesar 78.125%.

4. Kesimpulan

Berdasarkan penelitian mengenai implementasi *K-means clustering* pada aplikasi *GIS* didapat kesimpulan yang diperoleh setelah melakukan penelitian adalah sebagai berikut :

1. Pada metode *K-means clustering* yang telah diimplementasikan pada pengelompokkan hasil pertanian telah berhasil diimplementasikan pada *GIS (Geographic Information System)* sehingga mendapatkan pengelompokkan daerah sukses panen, kurang tercapai dan gagal panen.
2. Pengujian akurasi *k-means* yang sudah diujicobakan dengan 16 *dataset* dari tahun 2012 – 2015 didapatkan hasil akurasi yaitu 78.125%.
3. Pada proses penentuan centroid awal dilakukan secara acak, hanya saja ketika menentukan centroid harus melihat data riilnya terlebih dahulu sehingga waktu proses dan iterasi akan semakin sedikit.

Daftar Pustaka

- [1] Ilyas, Husni.2008. *Algoritma K-Means Clustering* (Diakses 23 Maret 2016) <https://komputasi.wordpress.com/2008/11/14/algoritma-clustering-k-means/>
- [2] Felicia, Lianna. *Penerapan Metode Clustering Dengan K-Means Untuk Memetakan Potensi Tanaman Padi Di Kota Semarang*. Semarang.
- [3] Utama, Integra. 2014. *GIS (Geographic Information System)* (Diakses 23 Maret 2016) <http://integrasiutama.com/gis-geographic-information-system/>

- [4] Kenneth E. Foote and Margaret Lynch , "*Geographic Information Systems as an Integrating Technology: Context, Concepts, and Definitions*"., *The Geographer's Craft Project, Department of Geography, The University of Colorado at Boulder*. Retrieved.
- [5] Kania,febby.2013. "*Sistem Informasi Geografi*"
- [6] BenZone.2013. "*Analisis Cluster*" <http://bbheche.blogspot.co.id/2013/07/analisis-cluster.html>