

KLASTERISASI DATA PERTANIAN DI KABUPATEN LAMONGAN MENGGUNAKAN ALGORITMA K-MEANS DAN FUZZY C MEANS

Arif Rohmatullah¹⁾, Dinita Rahmalia²⁾, Mohammad Syaiful Pradana³⁾

^{1) 2) 3)} Fakultas MIPA, Prodi Matematika, Universitas Islam Darul Ulum Lamongan, Jl. Airlangga No. 3 Lamongan;
Email: ¹⁾arifrohmatullah@unisda.ac.id, ²⁾dinitarahmalia@gmail.com, ³⁾syaifulp@unisda.ac.id

Abstrak

Di Indonesia terdapat beberapa orang yang bermatapencaharian sebagai petani untuk memenuhi kebutuhan pokok pangan serta karena memiliki lahan pertanian yang luas. Karena terdapat perbedaan luas lahan pertanian dan hasil produksi pertanian, maka diperlukan klusterisasi pada data pertanian. Tujuan klustering adalah untuk mengidentifikasi suatu kelompok data dari populasi data untuk menghasilkan sifat-sifat dari data itu sendiri. Pada penelitian ini akan digunakan dua metode yaitu : algoritma K-Means dan algoritma Fuzzy C Means (FCM). Algoritma K-Means dan algoritma FCM dapat mengkluster beberapa kecamatan di kabupaten Lamongan berdasarkan luas lahan pertanian dan hasil produksi pertanian. Pada algoritma K-Means, titik pusat kluster diupdate sehingga menghasilkan jumlahan euclidean distance yang minimum. Pada algoritma FCM, derajat keanggotaan (*the degree of membership*) diupdate sehingga menghasilkan nilai fungsi objective yang minimum. Berdasarkan hasil simulasi, kedua metode tersebut dapat mengkluster beberapa kecamatan di Kabupaten Lamongan berdasarkan luas lahan pertanian dan hasil produksi pertanian.

Kata kunci: Klustering, K-Means, Fuzzy C Means.

Abstract

In Indonesia, there are many farmers as livelihood because the demand of food and large field for agriculture. Due to the difference of agriculture area and agriculture production, then clustering is required. The purpose of clustering is to identify natural groupings of data from a large data set to produce a concise representation of a system behavior. In this research, two methods will be applied : K-Means algorithm and Fuzzy C Means (FCM) algorithm. Both K-Means algorithm and FCM algorithm can cluster some regions in Lamongan based on the agriculture area and agriculture production. For K-Means algorithm, we update cluster center so that minimizing the sum of optimum euclidean distance. For FCM, we update degrees of membership so that minimizing the objective function value. From the simulations, we can cluster some regions in Lamongan based on the agriculture area and agriculture production.

Keywords: Clustering, K-Means, Fuzzy C Means

1. PENDAHULUAN

Di Indonesia, terdapat beberapa pekerja sebagai petani sebagai matapencaharian karena kebutuhan pokok pangan. Lamongan adalah salah satu kabupaten di Jawa Timur, Indonesia yang memiliki lahan yang luas untuk pertanian. Berdasarkan data pertanian pada beberapa kecamatan di kabupaten Lamongan, terdapat perbedaan pada data luas lahan pertanian dan hasil produksi pertanian. Karena perbedaan tersebut maka pemerintah atau dinas di kabupaten Lamongan harus membuat klusterisasi data pertanian pada beberapa kecamatan di kabupaten Lamongan untuk memberikan informasi.

Klustering adalah proses mengelompokkan kumpulan object data ke dalam satu atau lebih kelompok sehingga data yang terhimpun pada suatu kelompok memiliki tingkat kesamaan (similarity) yang tinggi. Tujuan klustering adalah untuk mengidentifikasi suatu kelompok data dari populasi data untuk menghasilkan sifat-sifat dari data itu sendiri. Pada penelitian

sebelumnya, algoritma klustering seperti Kohonen Network (Rahmalia dkk, 2018) dan K-Medoids (Naimah dkk, 2017) telah digunakan. Pada penelitian ini akan digunakan dua metode yaitu : algoritma K-Means dan algoritma Fuzzy C Means (FCM).

Algoritma K-Means adalah salah satu algoritma klustering yang sederhana. Algoritma K-Means adalah algoritma pengklasteran berdasarkan nilai titik centroid. Titik centroid dari suatu kluster adalah titik pusat kluster.

Sedangkan himpunan Fuzzy diperkenalkan oleh L.A. Zadeh pada tahun 1965. Himpunan fuzzy berbeda dengan himpunan crisp dimana himpunan fuzzy memiliki fungsi keanggotaan. Terdapat beberapa aplikasi pada himpunan dan sistem fuzzy yaitu optimisasi (Rao, 2009), pengambilan keputusan (Kusumadewi dkk, 2006), kontrol (Herrera dkk, 2013), dan peramalan data (Dewi dkk, 2014). Aplikasi himpunan fuzzy juga dapat digunakan untuk masalah klusterisasi yaitu Fuzzy C Means (Zadeh, 1965).

Fuzzy C Means (FCM) adalah metode klustering data dimana setiap data masuk ke dalam kluster yang ditentukan oleh derajat keanggotaan (the degree of membership). Metode ini ditemukan oleh Bezdek pada tahun 1981 (Bezdek, 1981).

Algoritma K-Means dan algoritma FCM dapat mengkluster beberapa kecamatan di kabupaten Lamongan berdasarkan luas lahan pertanian dan hasil produksi pertanian. Pada algoritma K-Means, titik pusat kluster diupdate sehingga menghasilkan jumlahan euclidean distance yang minimum. Pada algoritma FCM, derajat keanggotaan (the degree of membership) diupdate sehingga menghasilkan nilai fungsi objective yang minimum. Berdasarkan hasil simulasi, kedua metode tersebut dapat mengkluster beberapa kecamatan di kabupaten Lamongan berdasarkan luas lahan pertanian dan hasil produksi pertanian.

2. METODE

Klustering adalah proses mengelompokkan kumpulan object data ke dalam satu atau lebih kelompok sehingga data yang terhimpun pada suatu kelompok memiliki tingkat kesamaan (similarity) yang tinggi. Klustering merupakan proses dari pengolahan data yang memiliki beberapa aplikasi. Tujuan klustering adalah untuk mengidentifikasi suatu kelompok data dari populasi data untuk menghasilkan sifat-sifat dari data itu sendiri (Han et al, 2012; Zaki and Meira, 2014). Pada penelitian ini akan digunakan dua metode untuk klustering yaitu : K-Means dan Fuzzy C-Means.

Algoritma K-Means

Algoritma K-Means adalah algoritma pengklasteran berdasarkan nilai titik centroid. Titik centroid dari suatu kluster adalah titik pusat kluster. Algoritma K-Means dapat dirancang sebagai berikut (Han, et al, 2012) :

Misalkan N adalah jumlah titik pusat kluster dan D adalah jumlah data.

1. Pilih N object data dari $\{x_1, x_2, \dots, x_D\}$ sebagai titik pusat kluster awal the $c_j, j = 1, 2, \dots, N$
2. Hitung euclidean distance dari data $x_i, i = 1, 2, \dots, D$ menuju titik pusat kluster $c_j, j = 1, 2, \dots, N$ menggunakan persamaan (1)

$$dist(i, j) = \|x_i - c_j\|, \quad (1)$$

$i = 1, 2, \dots, D, j = 1, 2, \dots, N$

3. Tentukan jarak minimum menuju titik pusat kluster untuk setiap data

$$distmin(i) = \|x_i - c^*\|, \quad (2)$$

$i = 1, 2, \dots, D$

4. Hitung fungsi objective menggunakan persamaan (3)

$$J = \sum_{i=1}^D \|x_i - c^*\| \quad (3)$$

5. Hitung rata-rata dari data yang telah diklusterkan sebagai titik pusat kluster yang baru $c_j, j = 1, 2, \dots, N$

Ulangi langkah 2-5 sampai aturan pemberhentian terpenuhi. Aturan pemberhentian dapat berupa maksimum iterasi atau nilai error terkecil tercapai.

Algoritma Fuzzy C Means

Fuzzy C Means (FCM) adalah metode klustering data dimana setiap data masuk ke dalam kluster yang ditentukan oleh derajat keanggotaan (the degree of membership). Metode ini ditemukan oleh Bezdek pada tahun 1981. Metode ini juga dapat mengkluster data yang memiliki ukuran dimensi ke dalam sejumlah kluster yang berbeda.

Optimisasi dari FCM adalah meminimumkan fungsi objective pada persamaan (4) :

$$J_m = \sum_{i=1}^D \sum_{j=1}^N \mu_{ij}^m \|x_i - c_j\|^2 \quad (4)$$

Dengan :

D : jumlah data

N : jumlah titik pusat kluster

m : fuzzy partition matrix exponent untuk mengendalikan derajat fuzzy, dengan $m > 1$

x_i : data ke- i

c_j : titik pusat kluster ke- j

μ_{ij} : derajat keanggotaan dari x_i pada titik pusat kluster ke- j

Untuk setiap data $x_i, i = 1, 2, \dots, D$ jumlah dari nilai fungsi keanggotaan untuk semua kluster adalah sama dengan satu seperti pada persamaan (7).

$$\mu = \begin{bmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{1N} \\ \mu_{21} & \mu_{22} & \dots & \mu_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{D1} & \mu_{D2} & \dots & \mu_{DN} \end{bmatrix} \quad (5)$$

$$\mu_{ij} \sim U(0,1), \quad i = 1, 2, \dots, D, \quad j = 1, 2, \dots, N \quad (6)$$

$$\sum_{j=1}^N \mu_{ij} = 1, \quad i = 1, 2, \dots, D \quad (7)$$

Algoritma FCM dapat dirancang sebagai berikut (Han et al, 2012):

1. Inisialisasi nilai derajat keanggotaan titik pusat kluster μ_{ij} seperti pada persamaan (6) dan persamaan (7) secara acak.
2. Hitung titik pusat kluster seperti pada persamaan (8) dengan $m = 2$.

$$c_j = \frac{\sum_{i=1}^D \mu_{ij}^m x_i}{\sum_{i=1}^D \mu_{ij}^m} \quad j = 1, 2, \dots, N \quad (8)$$

3. Update nilai μ_{ij} berdasarkan persamaan (9).

$$\mu_{ij} = \frac{1}{\sum_{k=1}^N \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (9)$$

$i = 1, 2, \dots, D \quad j = 1, 2, \dots, N$

4. Hitung fungsi objective J_m menggunakan persamaan (4).

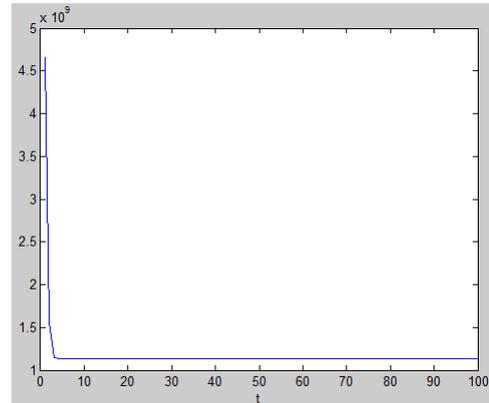
Ulangi langkah 2 – 4 sampai aturan pemberhentian terpenuhi. Aturan pemberhentian dapat berupa maksimum iterasi atau nilai error terkecil tercapai.

3. HASIL DAN PEMBAHASAN

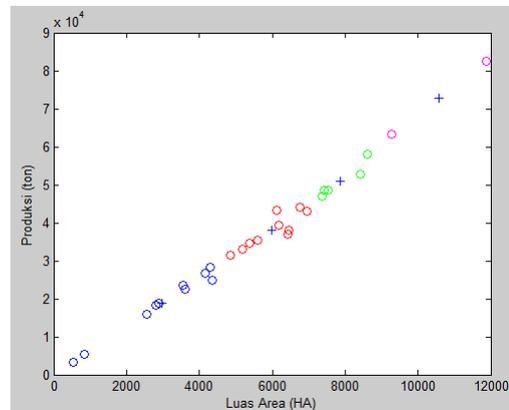
Data yang digunakan pada simulasi adalah data pertanian yang diperoleh dari Badan Pusat Statistik (BPS) kabupaten Lamongan yang terdiri dari : luas lahan pertanian (ha) dan hasil produksi pertanian (ton) pada 27 kecamatan di kabupaten Lamongan selama tiga tahun : 2013, 2014, and 2015. Untuk masing-masing tahun, simulasi menggunakan K-Means dan simulasi menggunakan FCM diterapkan untuk klusterisasi kecamatan di kabupaten Lamongan berdasarkan luas lahan pertanian dan hasil produksi pertanian.

Hasil Simulasi Algoritma K-Means

Pada data pertanian tahun 2013, hasil simulasi ditunjukkan pada Gambar 1 dan Gambar 2.



Gambar 1. Hasil Optimisasi dari Algoritma K-Means pada Data Pertanian Tahun 2013



Gambar 2. Hasil Klusterisasi dengan Empat Kluster pada Data Pertanian Tahun 2013

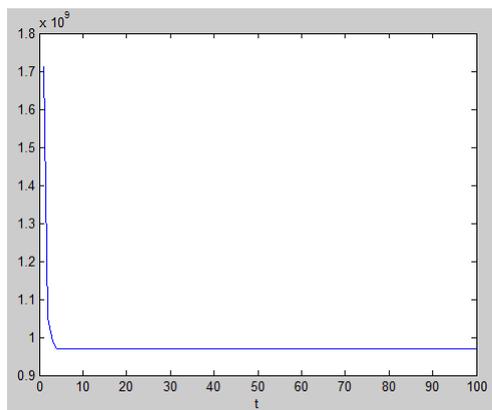
Gambar 1 adalah proses optimisasi dari algoritma K-Means. Pada proses optimisasi, jumlah euclidean distance konvergen menuju minimum nilai objective. Pada iterasi maksimum, nilai objective adalah 1128678845. Gambar 2 adalah hasil klusterisasi dengan empat kluster dengan tanda '+' adalah titik pusat kluster.

Dengan kecamatan yang telah terklusterisasi dapat dilihat pada Tabel 1.

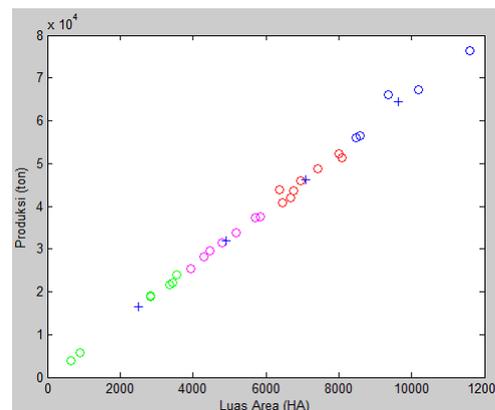
Pada data pertanian tahun 2014, hasil simulasi ditunjukkan pada Gambar 3 dan Gambar 4. Gambar 3 adalah proses optimisasi dari algoritma K-Means. Pada proses optimisasi, jumlah euclidean distance konvergen menuju minimum nilai objective. Pada iterasi maksimum, nilai objective adalah 970479764. Gambar 4 adalah hasil klusterisasi dengan empat kluster dengan tanda '+' adalah titik pusat kluster

Tabel 1. Hasil Klusterisasi Kecamatan Berdasarkan Data Pertanian Tahun 2013

Klaster 1	Klaster 2	Klaster 3	Klaster 4
3 Brondong	2 Bluluk	1 Babat	15 Modo
4 Deket	11 Lamongan	9 Kedungpring	23 Sugio
5 Glagah	12 Laren	10 Kembangbahu	
6 Kalitengah	14 Mantup	24 Sukodadi	
7 Karangbinangung	16 Ngimbang	26 Tikung	
8 Karanggeneng	18 Pucuk		
13 Maduran	19 Sambeng		
17 Paciran	20 Sarirejo		
22 Solokuro	21 Sekaran		
25 Sukorame	27 Turi		



Gambar 3. Hasil Optimisasi dari Algoritma K-Means pada Data Pertanian Tahun 2014



Gambar 4. Hasil Klasterisasi dengan Empat Klaster pada Data Pertanian Tahun 2014

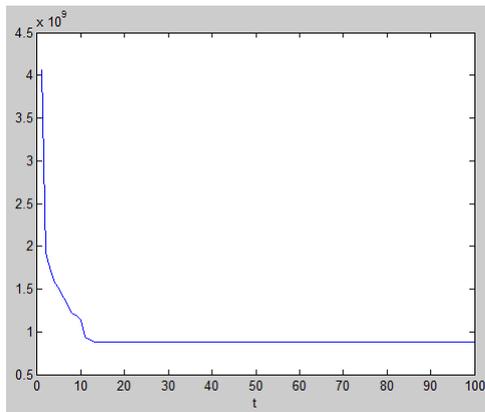
Dengan kecamatan yang telah terklasterisasi dapat dilihat pada Tabel 2.

Tabel 2. Hasil Klasterisasi Kecamatan Berdasarkan Data Pertanian Tahun 2014

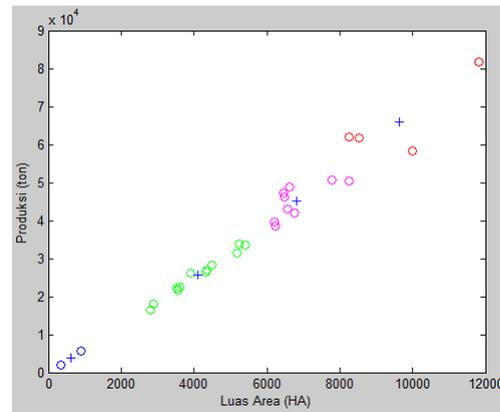
Klaster 1	Klaster 2	Klaster 3	Klaster 4
1 Babat	10 Kembangbahu	3 Brondong	2 Bluluk
9 Kedungpring	11 Lamongan	5 Glagah	4 Deket
12 Laren	16 Ngimbang	6 Kalitengah	7 Karangbinangung
15 Modo	18 Pucuk	8 Karanggeneng	13 Maduran
23 Sugio	20 Sarirejo	17 Paciran	14 Mantup
	21 Sekaran	22 Solokuro	19 Sambeng
	24 Sukodadi	25 Sukorame	27 Turi
	26 Tikung		

Pada data pertanian tahun 2015, hasil simulasi ditunjukkan pada Gambar 5 dan Gambar 6. Gambar 5 adalah proses optimisasi dari algoritma K-Means. Pada proses optimisasi, jumlah euclidean distance

konvergen menuju minimum nilai objective. Pada iterasi maksimum, nilai objective adalah 880641929. Gambar 6 adalah hasil klasterisasi dengan empat klaster dengan tanda '+' adalah titik pusat klaster.



Gambar 5. Hasil Optimisasi dari Algoritma K-Means pada Data Pertanian Tahun 2015



Gambar 6. Hasil Klasterisasi dengan Empat Klaster pada Data Pertanian Tahun 2015

Dengan kecamatan yang telah terklasterisasi dapat dilihat pada Tabel 3.

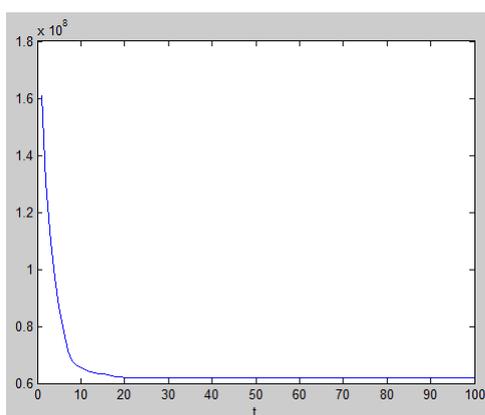
Tabel 3. Hasil Klasterisasi Kecamatan Berdasarkan Data Pertanian Tahun 2015

Klaster 1	Klaster 2	Klaster 3	Klaster 4
3 Brondong	9 Kedungpring	2 Bluluk	1 Babat
17 Paciran	15 Modo	4 Deket	10 Kembangbahu
	23 Sugio	5 Glagah	11 Lamongan
	24 Sukodadi	6 Kalitengah	12 Laren
		7 Karangbinangung	16 Ngimbang
		8 Karanggeneng	18 Pucuk
		13 Maduran	20 Sarirejo
		14 Mantup	21 Sekaran
		19 Sambeng	26 Tikung
		22 Solokuro	
		25 Sukorame	
		27 Turi	

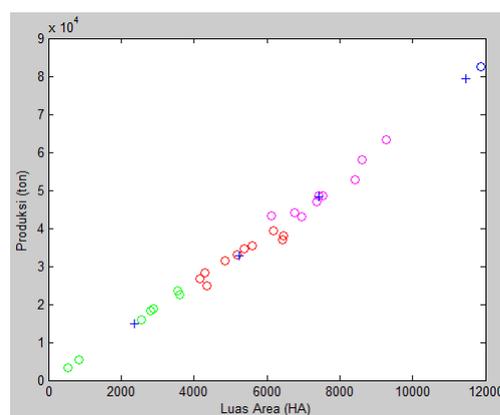
Hasil Simulasi Algoritma Fuzzy C Means

Pada data pertanian tahun 2013, hasil simulasi ditunjukkan pada Gambar 7 dan Gambar 8. Gambar 7 adalah proses optimisasi dari algoritma FCM. Pada proses optimisasi, jumlahan fungsi objective konver-

gen menuju minimum nilai objective. Pada iterasi maksimum, nilai fungsi objective adalah 61884425. Gambar 8 adalah hasil klasterisasi dengan empat klaster dengan tanda '+' adalah titik pusat klaster.



Gambar 7. Hasil Optimisasi dari Algoritma FCM pada Data Pertanian Tahun 2013



Gambar 8. Hasil Klasterisasi dengan Empat Klaster pada Data Pertanian Tahun 2013

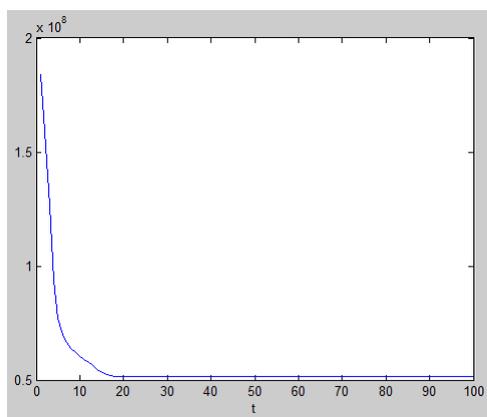
Dengan kecamatan yang telah terklasterisasi dapat dilihat pada Tabel 4.

Tabel 4. Hasil Klasterisasi Kecamatan Berdasarkan Data Pertanian Tahun 2013

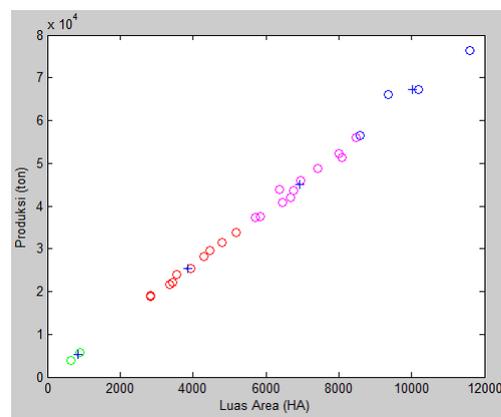
Klaster 1	Klaster 2	Klaster 3	Klaster 4
23 Sugio	2 Bluluk	3 Brodong	1 Babat
	4 Deket	5 Glagah	9 Kedungpring
	7 Karangbinangun	6 Kalitengah	10 Kembangbahu
	11 Lamongan	8 Karanggeneng	12 Laren
	13 Maduran	17 Paciran	15 Modo
	14 Mantup	22 Solokuro	21 Sarirejo
	16 Ngimbang	25 Sukorame	20 Sekaran
	18 Pucuk		24 Sukodadi
	19 Sambeng		26 Tikung
	27 Turi		

Pada data pertanian tahun 2014, hasil simulasi ditunjukkan pada Gambar 9 dan Gambar 10. Gambar 9 adalah proses optimisasi dari algoritma FCM. Pada proses optimisasi, jumlahan fungsi objective konver-

gen menuju minimum nilai objective. Pada iterasi maksimum, nilai fungsi objective adalah 51606178. Gambar 10 adalah hasil klasterisasi dengan empat klaster dengan tanda '+' adalah titik pusat klaster.



Gambar 9. Hasil Optimisasi dari Algoritma FCM pada Data Pertanian Tahun 2014



Gambar 10. Hasil Klasterisasi dengan Empat Klaster pada Data Pertanian Tahun 2014

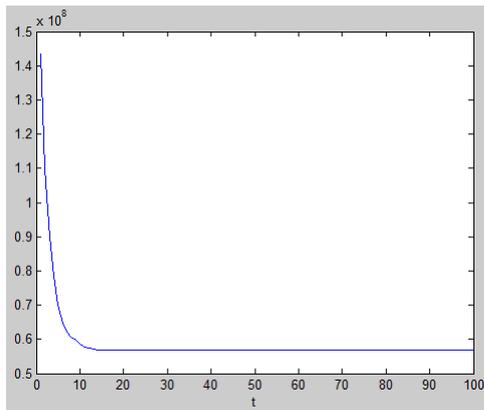
Dengan kecamatan yang telah terklasterisasi dapat dilihat pada Tabel 5.

Tabel 5. Hasil Klasterisasi Kecamatan Berdasarkan Data Pertanian Tahun 2014

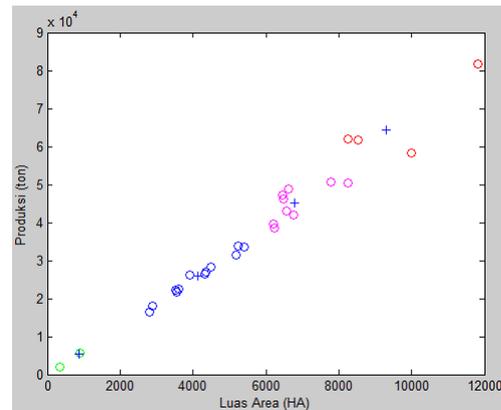
Klaster 1	Klaster 2	Klaster 3	Klaster 4
1 Babat	2 Bluluk	3 Brondong	10 Kembangbahu
9 Kedungpring	4 Deket	17 Paciran	11 Lamongan
15 Modo	5 Glagah		12 Laren
23 Sugio	6 Kalitengah		14 Mantup
	7 Karangbinangun		16 Ngimbang
	8 Karanggeneng		18 Pucuk
	13 Maduran		20 Sarirejo
	19 Sambeng		21 Sekaran
	22 Solokuro		24 Sukodadi
	25 Sukorame		26 Tikung
			27 Turi

Pada data pertanian tahun 2015, hasil simulasi ditunjukkan pada Gambar 11 dan Gambar 12. Gambar 11 adalah proses optimisasi dari algoritma FCM. Pada proses optimisasi, jumlahan fungsi objective konver-

gen menuju minimum nilai objective. Pada iterasi maksimum, nilai fungsi objective adalah 56757498. Gambar 12 adalah hasil klasterisasi dengan empat klaster dengan tanda '+' adalah titik pusat klaster.



Gambar 11. Hasil Optimisasi dari Algoritma FCM pada Data Pertanian Tahun 2015



Gambar 12. Hasil Klasterisasi dengan Empat Klaster pada Data Pertanian Tahun 2015

Dengan kecamatan yang telah terklasterisasi dapat dilihat pada Tabel 6.

Tabel 6. Hasil Klasterisasi Kecamatan Berdasarkan Data Pertanian Tahun 2015

Klaster 1	Klaster 2	Klaster 3	Klaster 4
2 Bluluk	9 Kedungpring	3 Brodong	1 Babat
4 Deket	15 Modo	17 Paciran	10 Kembangbahu
5 Glagah	23 Sugio		11 Lamongan
6 Kalitengah	24 Sukodadi		12 Laren
7 Karangbinangun			16 Ngimbang
8 Karanggeneng			18 Pucuk
13 Maduran			20 Sarirejo
14 Mantup			21 Sekaran
19 Sambeng			26 Tikung
22 Solokuro			
25 Sukorame			
27 Turi			

4. SIMPULAN

Algoritma K-Means dan FCM dapat mengklaster beberapa kecamatan di kabupaten Lamongan berdasarkan luas lahan pertanian dan hasil produksi pertanian. Pada algoritma K-Means, titik pusat klaster diupdate sehingga menghasilkan jumlahan euclidean distance yang minimum. Pada algoritma FCM, derajat keanggotaan (the degree of membership) diupdate sehingga menghasilkan nilai fungsi objective yang minimum. Berdasarkan hasil simulasi, beberapa kecamatan di kabupaten Lamongan dapat diklaster. Hasil klasterisasi beberapa kecamatan di kabupaten Lamongan dapat memberikan informasi mengenai potensi pertanian di kecamatan tersebut. Penelitian lebih lanjut dapat dikembangkan menggunakan data dengan dimensi yang lebih banyak.

5. UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan pada Kementerian Riset Teknologi dan Pendidikan Tinggi atas dana Penelitian Dosen Pemula tahun 2019.

6. DAFTAR PUSTAKA

- Bezdec, J.C. 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press
- Dewi, C., Kartikasari, D.P., Mursityo, Y.T. 2014. Prediksi Cuaca pada Data Time Series Menggunakan Adaptive Neuro Fuzzy Inference System (ANFIS). Jurnal Teknologi Informasi dan Ilmu Komputer. Vol. 1 No. 1
- Han, J., Kamber, M., Pei, J. 2012. Data Mining Concept and Techniques. Elsevier
- Herrera, et.al. 2013. Tuning of a TS Fuzzy Output Regulator Using the Steepest Descent Approach and ANFIS. Mathematical Problem in Engineering. Hindawi Publishing Corporation
- Kusumadewi, S., Hartati, S., Harjoko, A., Wardoyo, R., 2006. Fuzzy Multi Attribute Decision Making (Fuzzy MADM). Graha Ilmu, Yogyakarta.
- Naimah, E.N., Pradana, M.S., Amiroch, S. 2017. Penerapan Metode Clustering dengan K-Medoids untuk Memetakan Potensi Tanaman Padi di Kabupaten Lamongan. Skripsi Prodi Matematika Universitas Islam Darul Ulum Lamongan
- Rahmalia, D., Herlambang, T. 2018. Application Kohonen Network and Fuzzy C Means for Clustering Airports Based on Frequency of Flight. Kinetik : Game Technology, Information System, Computer Network. Vol. 3 No. 3

Rao, S.S. 2009. Engineering Optimization Theory and Practice. John Wiley and Sons, New Jersey
Zadeh, L.A. 1965. Fuzzy Set. Information and Control, vol. 8, pp. 338–353
Zaki, M.J., Meira, W. 2014. Data Mining and Analysis.

Fundamental Concepts and Algorithms.
Cambridge University Press
Zimmermann, H.J. 2001. Fuzzy Set Theory and Its Applications. Springer Science+Business Media