



Available online at : <http://bit.ly/InfoTekJar>

InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan

ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600



Implementasi Algoritma K-Means Dalam Penentuan Prioritas Rehabilitasi Daerah Aliran Sungai Cipunagara

Odi Nurdiawan¹, Fidya Arie Pratama²

¹ Dosen Program Studi Manajemen Informatika, STMIK IKMI Cirebon, Indonesia.

² Dosen Program Studi Komputerisasi Akuntansi, STMIK IKMI Cirebon, Indonesia

KEYWORDS

Daerah Aliran Sungai, Algoritma K-Means Clustering, Lahan Kritis

CORRESPONDENCE

Phone: 085 868 991 405

E-mail: odynurdiawan@gmail.com

ABSTRACT

Kekeringan salah satu kejadian yang sering terjadi pada wilayah indramayu bagian barat, rehabilitasi daerah aliran sungai menjadi salah satu alternatif. Lahan lahan wilayah subang - indramayu dalam keadaan kritis, maka perlu dilakukan pengelompokan wilayah lahan kritis menggunakan algoritma k-means, sehingga mudah dalam menentukan prioritas rehabilitasi daerah aliran sungai. Penelitian ini dibagi menjadi 3 tahap diantaranya Tahap 1 (Satu) pada tahap penelitian ini menentukan objek penelitian, yang akan dijadikan objek yaitu kelompok budidaya pertanian dan Bina Pengelola Daerah Aliran Sungai, kemudian dilakukan rumusan masalah. Tahap 2 (Dua) mengumpulkan data primer yang akan dilakukan *pre-processing*. Tahap 3 (Tiga) penerapan algoritma k-means dan Parameter data yang berpengaruh dalam menentukan tingkat kekritisian lahan yaitu skor penutupan lahan, skor lereng, skor erosi, skor produktivitas dan skor manajemen. Sehingga diketahui kelompok wilayah lahan kritis yang tinggi. Hasil penelitian ini rehabilitasi daerah aliran sungai cipunagara dengan menggunakan metode algoritma k-means dapat menghasilkan kumpulan *cluster* dengan indeks *Davies-Bouldin* terkecil kumpulan cluster terbaik adalah *cluster_1* dengan nilai 1886.707.

PENDAHULUAN

Kekeringan merupakan salah satu masalah lingkungan hidup yang secara spesifik dihadapi daerah indramayu yang setiap tahun terjadi. Hal ini tentu saja mengganggu kegiatan pertanian masyarakat, terlebih diketahui bahwa potensi Kabupaten subang-indramayu banyak terletak pada hasil pertanian seperti padi. Ketersediaan sumber air salah satu kebutuhan, mengingat selain potensi unggulan pada bidang pertanian memerlukan air cukup banyak. Daerah aliran sungai cipunagara adalah salah satu alternatif sebagai sumber air tidak bisa manampung air dalam debit yang banyak, di beberapa daliran sudah mengalami kerusakan[8-9]

Berdasarkan identifikasi dan pemetaan lahan kritis di daerah aliran sungai cipunagara ditentukan berdasarkan parameter Perdirjen BPDASPS Nomor P.4/V-Set/2013 ialah wilayah daerah aliran sungai cipunagara dibagi menjadi 3 kawasan utama yaitu kawasan hutan lindung, kawasan budidaya pertanian dan kawasan lindung di luar hutan serta hutan produksi. Data kawasan pada tabel 1 menjelaskan bahwa Kawasan

Budidaya Pertanian memiliki luas 106.801,47 Hektar, Kawasan Hutan Lindung memiliki luas 7.153,24 Hektar, dan kawasan hutan produksi memiliki 19.687,35 Hektar serta Kawasan lindung diluar 2.226,44 Hektar. Maka dari data tersebut Kawasan Budidaya Pertanian sangat dominan di wilayah subang dan indramayu serta wilayah yang dialiri sungai cipunagara. Hal tersebut menjadi salah satu sumber air bagi petani [4-5]

Tabel 1 Kawasan DAS Cipunagara

No	Kawasan	Luas (Ha)
1	Budidaya Pertanian	106.801,47
2	Hutan Lindung	7.153,24
3	Hutan Produksi	19.687,35
4	Lindung diluar Kawasan	2.226,44
TOTAL		135.868,50

Tabel 2 Parameter Tingkat Lahan Kritis

No	Parameter Lahan kritis	Penentuan HL	BP	KLLH dan HP
1	Penutupan Lahan	√	-	√
2	Produktivitas	-	√	-

3	Lereng	√	√	√
4	Tingkat Bahaya Erosi	√	√	√
5	Manajemen	√	√	√
6	Erovisitas Hujan (R)	-	-	-
7	Erosi yang diperoleh (EDP)	-	-	-
8	Erodibilitas	-	-	-

Keterangan : HL = Hutan Lindung, BP = Budidaya Pertanian, KLLH = Kawasan Lindung diluar kawasan hutan, dan HP = Hutan Produksi

Salah satu solusi dalam menangani kekeringan di wilayah daerah aliran sungai cipunegara yaitu dilakukan rehabilitasi lahan kritis karena tingkat penutupan lahan, tingkat bahaya erosi dan lereng sangat tinggi. Hal tersebut berdampak pada pertanian di wilayah subang-indramayu. Dalam melakukan rehabilitas perlu dilakukan pengelompokan wilayah lahan kritis, sedang maupun baik. Pengelompokan dilakukan menggunakan algoritma k-means. Data yang akan diambil dari Balai Besar Wilayah Sungai Citarum dan difokuskan pada Bina Pengelolaan Daerah Aliran Sungai Cipunagara dan Bendungan Salam Darma.

STUDI PUSTAKA

Lahan Kritis

Lahan kritis merupakan lahan yang telah mengalami atau dalam proses kerusakan fisik, kimia, atau biologi yang akhirnya dapat membahayakan fungsi hidrologi, orologi, produksi pertanian, pemukiman, dan kehidupan sosial ekonomi dari daerah lingkungan pengaruhnya. Peningkatan jumlah penduduk sejalan pula dengan meningkatnya kebutuhan akan lahan peruntukan pemukiman bagi tempat tinggal manusia, industri, maupun lahan pertanian sebagai sarana pemenuhan kebutuhan pangan manusia [7-10]

Data Mining

Data mining sebagai suatu proses eksplorasi dan analisis secara otomatis maupun semi otomatis terhadap data dalam jumlah besar dengan tujuan menemukan pola atau aturan yang berarti [1]

Algoritma K-Means

Algoritma K -means menggunakan proses secara berulang-ulang untuk mendapatkan basis data cluster. Dibutuhkan jumlah cluster awal yang diinginkan sebagai masukan dan menghasilkan titik centroid akhir sebagai output. Metode K-means akan memilih pola k sebagai titik awal centroid Secara acak atau random. Jumlah iterasi untuk mencapai cluster Centroid akan dipengaruhi oleh calon Cluster centroid awal secara random. Sehingga didapat cara dalam pengembangan algoritma dengan menentukan centroid Cluster yang dilihat dari kepadatan data awal yang tinggi agar mendapatkan kinerja yang lebih tinggi [17]

Berikut ini langkah-langkah yang terdapat pada algoritma K-Means [15]

- a) Tentukan k sebagai jumlah cluster yang dibentuk
- b) Tentukan k centroid (titik pusat cluster) awal secara random Penentuan centroid awal dilakukan secara random/acak dari objek-objek yang tersedia sebanyak k cluster, kemudian untuk menghitung centroid cluster ke-i berikutnya, digunakan rumus sebagai berikut :

$$v = \frac{\sum_{i=1}^n x_i}{n} ; i = 1, 2, 3, \dots, n$$

Dimana :

v : centroid pada cluster xi : objek ke-i

n : banyaknya objek/jumlah objek yang menjadi anggota cluster

- c) Hitung jarak setiap objek ke masing-masing centroid dari masing- masing cluster. Untuk menghitung jarak antara objek dengan centroid dapat menggunakan Euclidian Distance.

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; i = 1, 2, 3, \dots, n$$

Dimana; xi : objek x ke-i

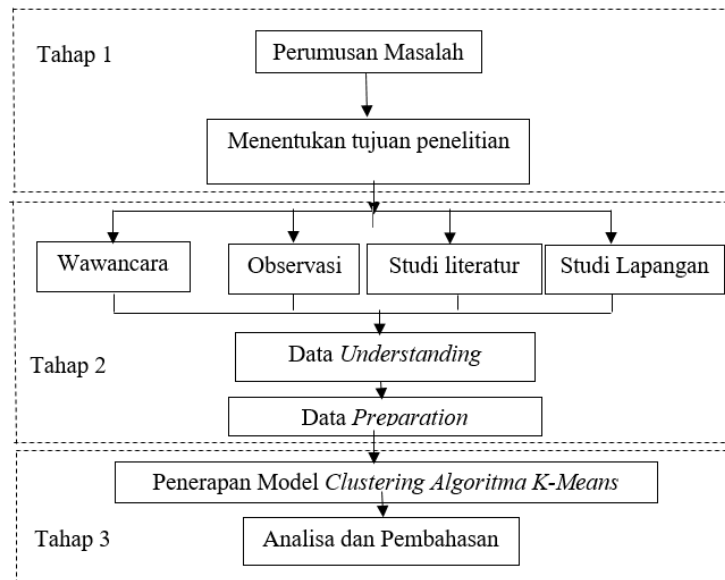
yi: daya y ke-i

n : banyaknya objek

- d) Alokasikan masing-masing objek ke dalam centroid yang paling dekat. Untuk melakukan pengalokasian objek kedalam masing-masing cluster pada saat iterasi secara umum dapat dilakukan dengan cara hard k- means dimana secara tegas setiap objek dinyatakan sebagai anggota cluster dengan mengukur jarak kedekatan sifatnya terhadap titik pusat cluster tersebut
- e) Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan
- f) Ulangi langkah 3 jika posisi centroid baru tidak sama

METODE PENELITIAN

Diagram Alir Penelitian Dapat dilihat pada gambar xx



Gambar 1 Diagram Alir

1. Tahap 1 (Satu)

Penelitian Awal memiliki beberapa tahapan di antaranya :

- a. Penentuan Objek Penelitian
- b. Rumusan Masalah
- c. Menentukan tujuan penelitian

2. Tahap 2 (Dua)

Pada tahap ini memiliki beberapa tahapan diantaranya :

- a. Observasi
 - b. Wawancara
 - c. Studi literatur
 - d. Studi Lapangan
 - e. Pemahaman Data (*Data Understanding*)
 - f. *Data Preparation*
3. **Tahap 3 (Tiga)**
Penerapan Model *Clustering Algoritma K-Means*

HASIL DAN PEMBAHASAN

Lokasi

Penelitian ini dilakukan di wilayah DAS Cipunagara yang sebagian besar berada di Kabupaten Subang dan indramayu. Data yang digunakan berupa peraturan Perundang undangan, BPDAS Citarum pada tahun 2016 dan 2017, dan instansi – instansi lain yang berkaitan dengan data yang diperlukan.

Populasi dan Sampling

Penelitian yang akan digunakan menggunakan data Balai Besar Wilayah Sungai Citarum difokuskan pada Badan Pengelolaan Daerah Aliran Sungai Cipunagara pada tahun 2016 dan 2017, yang mencakup 29 Kecamatan dan 207, menggunakan teknik sampling jenuh artinya sampel semua anggota populasi digunakan sebagai sampel. Sample yang digunakan ialah 414

Record data.

Tabel 3 Dataset DAS Cipunagara

No	Kec	DS	PL	KL	TBE	PO	MAN
1	Subang	Pegaden	77	25	95	90	Lengka p
2	Subang	Binong	60	25	95	90	Lengka p
3	Subang	Cipunagara	80	25	95	90	Lengka p
...
414	Anjatan	Anjatan	60	25	56	56	Lengka p

Keterangan :

- Kec : Kecamatan
- DS : Desa
- PL : Penutupan Lahan
- KL : Kemiringan Lahan
- TBE : Tingkat Bahaya Erosi
- PO : Produktivitas
- MAN : Manajemen

Preprocessing Data

Berdasarkan data set dilakukan pembobotan sesuai dengan Permenhut Nomor P.32/Menhut-II/2009.

Tabel 4 Pembobotan

NO	PL	KL	TBE	PO	MAN
1	4	2	95	5	5
2	3	2	95	5	5
3	4	2	95	5	5
...
414	3	2	95	5	5

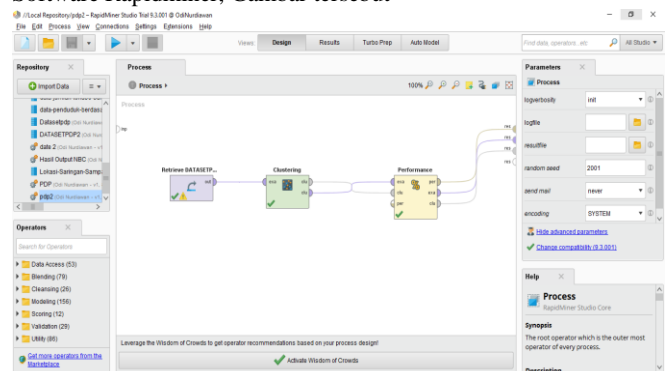
Hasil Normalisasi dari dataset, dapat dilihat pada tabel 5.

Tabel 5 Normalisasi

NO	PL	KL	TBE	PO	MAN
1	200	2	95	150	50
2	150	2	95	150	50
3	200	2	95	150	50
4	150	2	95	150	50
...
414	100	2	95	150	50

Penerapan Algoritma K-Means

Penerapan Algoritma K-means Menggunakan Software Rapidminer Versi 9. Langkah pertama ialah input data set dengan file spreadsheet (Microsoft Excel). Pembacaan file dataset tersebut menggunakan operator Read Exce yang ada pada Software Rapidminer, Gambar tersebut



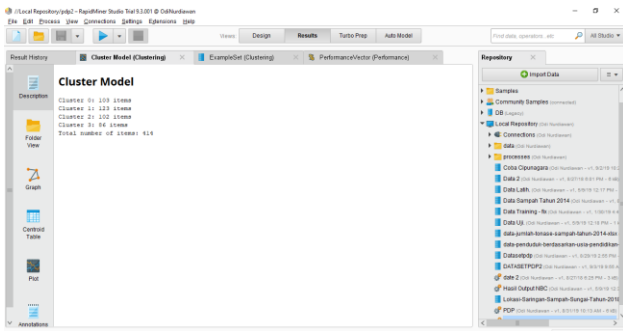
Gambar 2 Desain

Parameter penerapan algoritma k-means pada penelitian ini sebagai berikut :

Tabel 6 Parameter

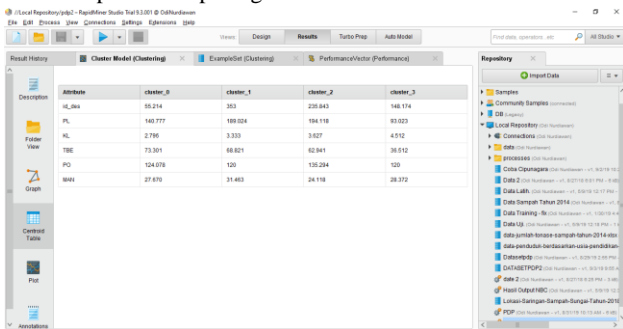
No	Parameter	Value
1	K	4
2	Max Run	10
3	Measure type	Numerical measurement
4	Numerical measure	EuclideanDistance
5	Max optimization step	100 (default)

Hasil Results model Algoritma K-means terhadap 4 (empat) Cluster data tersebut dapat dilihat pada gambar 3.



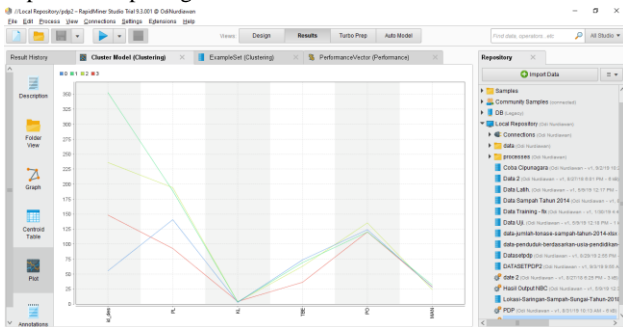
Gambar 3 Cluster Model

Hasil Centroid dengan 4 (empat) Cluster pada Cluster Model K-Means dapat dilihat pada gambar 4



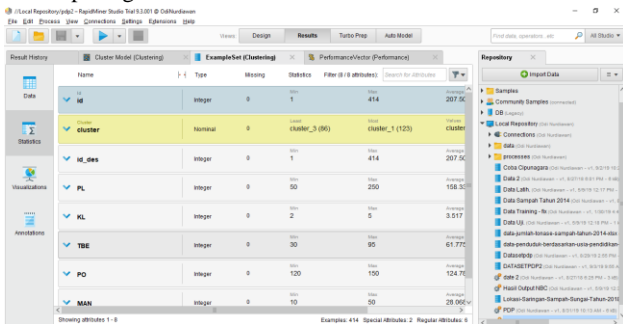
Gambar 4 Centroid

Hasil Plot dengan 4 (empat) Cluster pada Cluster Model K-Means dapat dilihat pada gambar 5



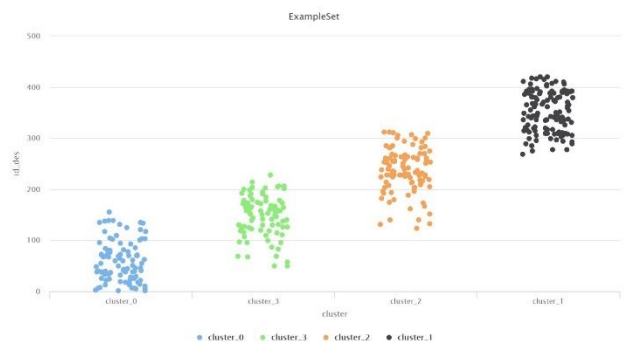
Gambar 5 Plot

Hasil Set Cluster berdasarkan Statistic dengan 6 (Enam) Atribut yang akan digunakan, diantaranya Id_desa, PL (Penutupan Lahan), KL (Kemiringan Lahan), TBE (Tingkat Bahaya Erosi), PO (Produktivitas), MAN (Manajemen), data set cluster dapat dilihat pada gambar 6



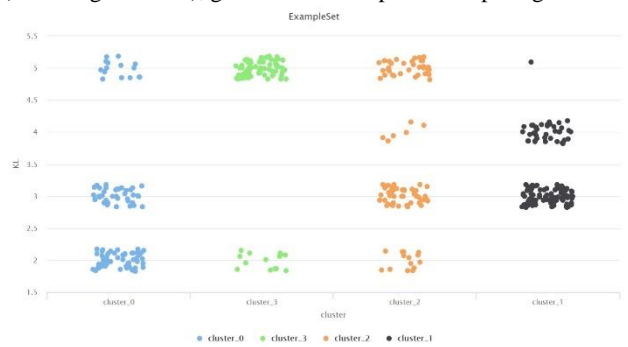
Gambar 6 Hasil Set Cluster

Hasil Set Cluster visualizations berdasarkan grafik scatter Id_Desa, grafik tersebut dapat dilihat pada gambar 7



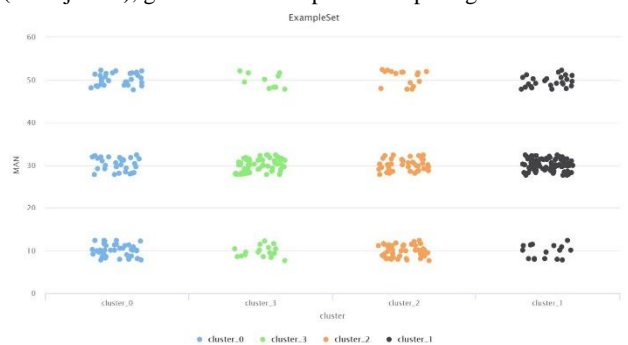
Gambar 7 Hasil Set Cluster

Hasil Set Cluster visualizations berdasarkan grafik scatter KL (Kemiringan Lahan), grafik tersebut dapat dilihat pada gambar 8



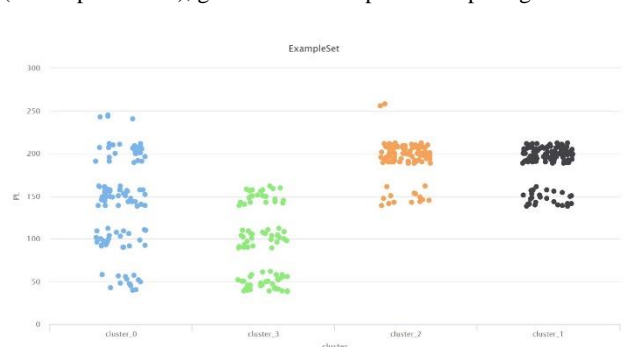
Gambar 8 Hasil Set Cluster

Hasil Set Cluster visualizations berdasarkan grafik scatter MAN (Manajemen), grafik tersebut dapat dilihat pada gambar 9



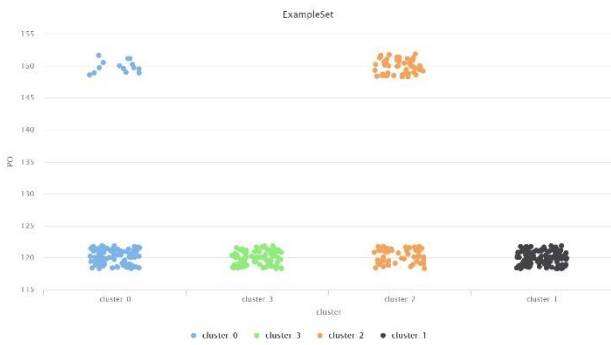
Gambar 9 Hasil Set Cluster

Hasil Set Cluster visualizations berdasarkan grafik scatter PL (Penutupan Lahan), grafik tersebut dapat dilihat pada gambar 10



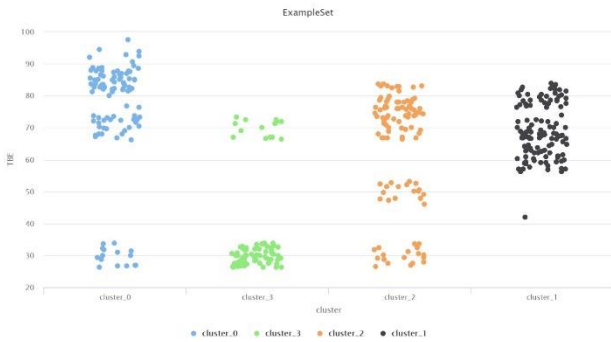
Gambar 10 Hasil Set Cluster

Hasil Set Cluster visualizations berdasarkan grafik scatter PO (Produktivitas), grafik tersebut dapat dilihat pada gambar 11



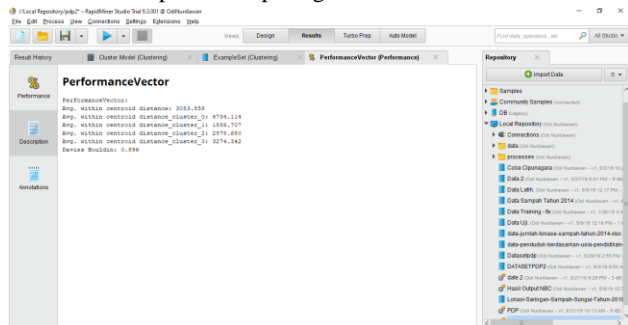
Gambar 11 Hasil Set Cluster

Hasil Set Cluster visualizations berdasarkan grafik scatter TBE (Tingkat Bahaya Erosi), grafik tersebut dapat dilihat pada gambar 12



Gambar 12 Hasil Set Cluster

Hasil Performance Vector berdasarkan rata rata setiap cluster, data tersebut dapat dilihat pada gambar 13



Gambar 13 Hasil Performance Vector

Pembahasan

Berdasarkan pemaparan diatas tersebut untuk mengevaluasi kinerja cluster menggunakan Operator *Cluster Distance Performance* mengambil model *cluster centroid* dan mengatur *cluster* sebagai input dan mengevaluasi kinerja model berdasarkan *centroid cluster* Operator ini digunakan untuk evaluasi kinerja metode pengelompokan berdasarkan centroid. Hasil dari *Cluster Distance Performance* menjelaskan bahwa *performance vector* pada *Avg.within centroid distance* sebesar 3058,858. *Avg.within centroid distance_cluster_0* sebesar 4734,114. *Avg.within centroid distance_cluster_1* sebesar 1886.707. *Avg.within centroid distance_cluster_2* sebesar 2578,680. *Avg.within centroid distance_cluster_3* sebesar 3274.342. sedangkan *Davies Bouldin* sebesar 0.896. maka *cluster* dengan jarak *intra-cluster* yang rendah (kesamaan *intra-cluster* yang tinggi) dan jarak antar-*cluster* yang tinggi (kesamaan antar-*cluster* yang rendah) akan memiliki indeks *Davies-Bouldin* yang

rendah, algoritma *clustering* yang menghasilkan kumpulan *cluster* dengan indeks *Davies-Bouldin* terkecil kumpulan cluster terbaik adalah *cluster_1* dengan nilai 1886.707.

Hasil *cluster* terbaik adalah *cluster_1* maka prioritas rehabilitasi daerah aliran sungai cipunagara, dapat dilihat pada tabel 7

Tabel 7 Cluster 1

No	No Id	Cluster 1	Nama Desa
1	292	cluster_1	Cisaat
2	293	cluster_1	Curugrendeng
3	294	cluster_1	Sarireja
4	295	cluster_1	Sanca
...
123	296	cluster_1	Cimanglid

KESIMPULAN

Hasil simpulan pada penelitian ini bahwa rehabilitasi daerah aliran sungai cipunagara dengan menggunakan metode algoritma k-means dapat mengevaluasi kinerja cluster menggunakan Operator *Cluster Distance Performance* mengambil model *cluster centroid* dan mengatur *cluster* sebagai input dan mengevaluasi kinerja model berdasarkan *centroid cluster* Operator ini digunakan untuk evaluasi kinerja metode pengelompokan berdasarkan centroid. Hasil dari *Cluster Distance Performance* menjelaskan bahwa *performance vector* pada *Avg.within centroid distance* sebesar 3058,858. *Avg.within centroid distance_cluster_0* sebesar 4734,114. *Avg.within centroid distance_cluster_1* sebesar 1886.707. *Avg.within centroid distance_cluster_2* sebesar 2578,680. *Avg.within centroid distance_cluster_3* sebesar 3274.342. sedangkan *Davies Bouldin* sebesar 0.896. maka *cluster* dengan jarak *intra-cluster* yang rendah (kesamaan *intra-cluster* yang tinggi) dan jarak antar-*cluster* yang tinggi (kesamaan antar-*cluster* yang rendah) akan memiliki indeks *Davies-Bouldin* yang rendah, algoritma *clustering* yang menghasilkan kumpulan *cluster* dengan indeks *Davies-Bouldin* terkecil kumpulan cluster terbaik adalah *cluster_1* dengan nilai 1886.707

ACKNOWLEDGMENT

Peneliti Mengucapkan Banyak Terimakasih kepada Kementerian Riset Teknologi dan Pendidikan Tinggi atas Hibah Penelitian Dosen Pemula, STMIK IKMI Cirebon, Balai Besar Wilayah Sungai Citarum, Bina Pengelolaan Daerah Aliran Sungai Cipunagara dan Bendungan Salam Darma Serta Kelompok Tani wilayah Gabus Wetan.

DAFTAR PUSTAKA

[1] Larose , Daniel T, 2005, Discovering Knowledge in Data: An Introduction to Data Mining, John Willey & Sons. Inc
 [2] Agusta, Y. 2007. K-means - Penerapan, Permasalahan dan Metode Terkait. Jurnal Sistem dan Informatika Vol. 3 (Februari 2007): 47-60.
 [3] Kementerian Kehutanan. 2013. Peraturan Direktur Jenderal Bina Pengelolaan Daerah lauran Sungai dan Perhutanan Sosial Nomor : P 4/V-Set/2013 tentang

- Petunjuk Teknis Penyusunan Data Spasial Lahan Kritis.
Jakarta (ID) : Kemenhut
- [4] Peraturan Direktur Jendral Bina Pengelolaan Daerah Aliran Sungai dan Perhutanan Sosial Nomor. P.4/V-Set/2013. Petunjuk Teknis Penyusunan Data Spasial Lahan Kritis. Jakarta: BPDASPS
- [5] Peraturan Menteri Kehutanan Nomor. P.32/Menhut-II/2009. Petunjuk Teknis Penyusunan Data Spasial Lahan Kritis. Jakarta: Kementrian Kehutanan.
- [6] Poerwowidodo. 1990. Telaah Kesuburan Tanah. Bamdung: Angkasa
- [7] Hanifah, K. A., 2012. Dasar-dasar Ilmu Tanah. Penerbit PT Raja Grafindo Persada. Jakarta.
- [8] Arsad, Sitanala. 1989. Konservasi Tanah dan Air. Bogor: IPB Press
- [9] Hardjowigeno, S. 2007. Ilmu Tanah. Jakarta: Akademika Pressindo. 296 Halaman
- [10] Hardjowigeno, S., Subagyo, H., dan Luthfi, R.M. 2004. Morfologi dan Klasifikasi Tanah Sawah. Di dalam: Tanah Sawah dan Teknologi pengelolaannya. Pusat Penelitian Tanah dan Agroklimat. Departemen Pertanian : Bogor.
- [11] Hardjowigeno, S. dan M. L. Rayes. 2005. Tanah Sawah Karakteristik, Kondisi dan Permasalahan Tanah Sawah di Indonesia. Bayumedia Publishing. Malang.
- [12] Sunarto. 1991. Geomorfologi Pantai. Pusat Antar Universitas Ilmu Teknik UGM. Yogyakarta
- [13] Sarief, S. 1980. Fisika Tanah Dasar. Serial Publikasi Ilmu-ilmu Tanah. Fakultas Pertanian. Universitas Padjajaran: Bandung. 120 hal.
- [14] Utomo, W.H. 1989. Erosi dan Konservasi Tanah, IKIP Malang, Malang.
- [15] Turban, Efraim, et al. 2005. Decision Support Systems and Intelligent Systems 7th Ed. New Jersey : Pearson Education
- [16] Madhulatha, T.S., 2012. AN OVERVIEW ON CLUSTERING METHODS. IOSR Journal of Engineering, II(1), pp.719-25
- [17] HUNG, C.M., WU, J., CHANG, J.H. & YANG, D.L., 2005. An Efficient k-Means Clustering Algorithm Using Simple Partitioning. JOURNAL OF INFORMATION SCIENCE AND ENGINEERING, XXI(1), pp.1157-77.
- [18] Bataineh, K.M., Naji, M. & Saqer, M., 2011. A Comparasion Study between Various Fuzzy Clustering Algorithms. Jordan Journal of Mechanical and Industrial Engineering , V(4), pp.335-43.
- [19] Varghese, B.M., J, J.T., A, U. & K, J.P., 2011. Clustering Student Data to Characterize Performance Patterns. International Journal of Advanced Computer Science and Applications, I(3), pp.138-40.
- [20] Bangoria, B., Mankad, N. & Pambhar, V., 2013. A Survey on Efficient Enhanced K-Means Clustering Algorithm. International Journal for Scientific Research & Development, I(9), pp.1698-700.
- [21] Ediyanto, Mara, M.N. & Satyahadewi, N., 2013. Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. Buletin Ilmiah Mat. Stat. dan Terapannya , II(2), pp.133-36.