

PENERAPAN CLUSTERING DBSCAN UNTUK PERTANIAN PADI DI KABUPATEN KARAWANG

Betha Nurina Sari¹⁾, Aji Primajaya²⁾

^{1,2)} Teknik Informatika, Universitas Singaperbangsa Karawang

Jl. H.S. Ronggowaluyo Teluk Jambe Karawang 41361

e-mail : betha.nurina@staff.unsika.ac.id¹⁾, aji.primajaya@staff.unsika.ac.id²⁾

ABSTRAK

Kabupaten Karawang merupakan salah satu kabupaten dengan hasil produksi padi tertinggi di Indonesia untuk kebutuhan pangan provinsi dan nasional, sehingga sering disebut sebagai kota lumbung padi. Kabupaten Karawang terdiri dari 30 kecamatan, masing-masing kecamatan memiliki luas lahan sawah yang berbeda luasnya, sehingga berbeda pula potensi produksi padi sawah yang bisa dipanen. Pemetaan yang dilakukan agar bisa menganalisis terkait karakteristik dari setiap kategori yang dibagikan. algoritma clustering yang diterapkan dalam penelitian ini adalah DBSCAN (Density-Based Spatial Clustering). Tujuan dari penelitian ini adalah menerapkan teknik clustering DBSCAN pada lahan pertanian padi di kabupaten Karawang, melakukan perbandingan evaluasi performa dari teknik clustering DBSCAN yang dilakukan pada penelitian ini. Penerapan teknik clustering DBSCAN (Density-Based Spatial Clustering of Application with Noise) pada lahan pertanian padi di kabupaten Karawang melalui beberapa tahapan penelitian, yaitu preprocessing, data mining, evaluasi cluster dan visualisasi melalui WEB-GIS. Algoritma DBSCAN diproses melalui R studio dengan bahasa pemrograman R dengan beberapa skenario eksperimen kombinasi masukan epsilon dan minPts. Perbandingan evaluasi performa dari teknik clustering DBSCAN (Density-Based Spatial Clustering of Application with Noise) dilakukan dengan memperhatikan nilai dari average silhouette width. Besarnya tingkat skor silhouette menunjukkan kualitas cluster yang terbentuk. Hasil eksperimen pada penelitian ini menunjukkan skor hasil tertinggi 0,74 dengan dua cluster.

Kata Kunci: clustering, DBSCAN, lahan, pertanian, Karawang

ABSTRACT

Karawang regency is one of the districts with the highest rice production in Indonesia for provincial and national food needs, so it is often referred to as the city of rice barns. karawang regency consists of 30 sub-districts, each sub-district has a wide area of rice fields that are different in width, so that the potential for production of lowland rice can be harvested. mapping is done so that it can analyze the characteristics of each category that is shared. the clustering algorithm applied in this research is dbscan (density-based spatial clustering). the purpose of this study was to apply dbscan clustering techniques on rice farming land in karawang regency, to do a comparison of performance evaluation from dbscan clustering techniques conducted in this study. application of dbscan (density-based spatial clustering of application with noise) clustering techniques on rice farmland in karawang regency through several stages of research, namely preprocessing, data mining, cluster evaluation and visualization through web-gis. the dbscan algorithm is processed through r studio with r programming language with several experimental scenarios in combination with epsilon and minpts inputs. comparison of performance evaluation from dbscan (density-based spatial clustering of application with noise) clustering techniques is done by considering the value of average silhouette width. the level of the silhouette score indicates the quality of the cluster formed. the experimental results in this study showed the highest yield score of 0.74 with two clusters.

Keywords: clustering, DBSCAN, lahan, pertanian, Karawang

I. PENDAHULUAN

Kabupaten Karawang merupakan salah satu kabupaten dengan hasil produksi padi tertinggi di Indonesia untuk kebutuhan pangan provinsi dan nasional, sehingga sering disebut sebagai kota lumbung padi. Sekitar 71,85% dari target yaitu 647.655 ton dari target 1.489.781 ton GKP produksi padi ini dicapai pada tahun 2015. Berdasarkan data dari Dinas Pertanian Kehutanan Perkebunan dan Peternakan Kab. Karawang, perkembangan produktivitas padi sawah mulai tahun 2010 sampai tahun 2015 mengalami perkembangan dari tahun ke tahun, ada kenaikan dan ada juga penurunan terkait tingkat produktivitas.

Kabupaten Karawang terdiri dari 30 kecamatan, masing-masing kecamatan memiliki luas lahan sawah yang berbeda luasnya, sehingga berbeda pula potensi produksi padi sawah yang bisa dipanen. Tidak hanya luas lahan sawah, luas lahan tanam dan panen setiap tahun juga mengalami perubahan dari tahun ke tahun, baik karena faktor alam (curah hujan, hama, tanah, bencana banjir) maupun faktor lain (pengalihan fungsi lahan, manusia), hal ini yang menyebabkan tingkat produktivitas juga berubah. Ada pembagian target produksi di masing-masing kecamatan untuk mendukung target capaian tingkat kabupaten, perlu ada pemetaan terkait kondisi yang berbeda dan strategi

penanganan dan optimalisasi yang berbeda pula. Penelitian ini bertujuan untuk pemetaan atau pengelompokan daerah sesuai karakteristiknya tersebut.

Teknik data mining yang dapat diterapkan untuk permasalahan tersebut adalah teknik *clustering*. Dalam data mining, clustering merupakan teknik *unsupervised learning*, yaitu pembelajaran tidak dibimbing atau disupervisi, pembelajaran tanpa data latih terlebih dahulu. Teknik clustering adalah pengelompokan data berdasarkan kemiripannya menjadi beberapa kelompok (*cluster*). Setiap *cluster* memiliki karakteristik khusus yang membedakan dengan cluster yang lain. Salah satu algoritma clustering yang diterapkan dalam penelitian ini adalah DBSCAN (*Density-Based Spatial Clustering*).

DBSCAN (*Density-Based Spatial Clustering Of Applications With Noise*) merupakan algoritma yang terpilih mendapatkan penghargaan the test of the time award pada tahun 2014 oleh ACM (*Association for Computing Machinery*) saat konferensi data mining. Algoritma DBSCAN dalam melakukan proses clustering untuk bentuk cluster yang tidak beraturan bisa lebih efisien dibandingkan algoritma CLARANS. Proses clustering yang dilakukan berdasarkan tingkat kedekatan/kepadatan jarak antar obyek dalam dataset sehingga masuk dalam kategori *density-based clustering*. Kelebihannya algoritma ini adalah dapat mendeteksi outlier/noise, tidak perlu mendapatkan input awalan berupa jumlah cluster (*k*) seperti k-Means atau k-Medoids, dan dapat mengenali bentuk cluster yang sulit tidak beraturan. Kelemahannya algoritma ini adalah terkait penentuan dua nilai input ϵ dan minPts karena mempengaruhi hasil clustering [1].

Penelitian dalam masalah clustering yang menggunakan DBSCAN juga dilakukan sebelumnya oleh Andreas Yuwono, dkk pada tahun 2014 untuk mencari arah penyebaran wabah demam berdarah di Yogyakarta [2]. Hasil penelitian menunjukkan bahwa metode DBSCAN dapat memberikan informasi terkait tingkat kepadatan penderita demam berdarah tinggi dan rendah. Pada tahun 2015 penelitian implementasi metode clustering DBSCAN pada proses pengambilan keputusan dilakukan untuk membantu perusahaan menentukan pelanggan potensialnya [3]. Selain itu ada pula penelitian yang implementasi Density Based Clustering menggunakan Graphics Processing Unit (GPU) [4]. Pada penelitian yang dilakukan Parasian menggunakan metode DBSCAN Clustering pada data penyakit pasien khususnya pengguna BPJS sebagai usaha untuk menemukan kecenderungan penyakit yang ada di sekelompok masyarakat yang berobat ke RSUP Haji Adam Malik Medan [5]. Muhammad Tanzil Furqon dan Lailil Muflikhah melakukan penelitian dengan judul *Clustering The Potential Risk Of Tsunami Using Density-Based Spatial Clustering Of Application With Noise* (DBSCAN) pada tahun 2016 [1], menunjukkan hasil analisis potensi resiko tsunami.

Model yang dihasilkan dari teknik clustering direpresentasikan dalam berbagai jenis bentuk, salah satunya berupa sistem informasi geografis berbasis web. Terkait dengan teknik representasi atau visualisasi *knowledge*, terdapat beberapa penelitian yang telah dilakukan, yaitu penelitian tentang visualisasi secara spasial cluster kerusakan sarana dan prasarana sekolah oleh Rudy Herteno pada tahun 2016 [6], Jingyuan Zhagn pada tahun 2014 meneliti tentang visualisasi web geospasial untuk analisis clustering pada data epimologikal [7]. Hasil penelitian ini berupa analisis pemetaan berdasarkan karakteristik dari penerapan clustering DBSCAN diharapkan dapat menjadi salah satu pendekatan dan analisis terkait kondisi lahan pertanian di kabupaten Karawang.

II. METODE

Objek penelitian ini adalah pengelompokan (*clustering*) kecamatan berdasarkan karakteristik yang ada di lahan pertanian kabupaten Karawang. Data yang digunakan pada penelitian ini adalah data laporan hasil pertanian di Kabupaten Karawang mulai tahun 2010 sampai tahun 2015 yang diambil dari Dinas Pertanian dan Kehutanan Kabupaten Karawang bagian pangan. Data terdiri dari 15 variabel, yaitu data luas sawah, luas baku sawah, luas tanam, luas panen, jumlah hasil produksi, curah hujan, hari hujan, organisme pengganggu tanaman (penggerek batang, tikus, wereng, batang cokelat, blasit, siput murbai, bakteri hawar daun, hama putih palsu), dan tingkat produktivitas padi.

Tahapan penelitian yang akan dilakukan adalah sebagai berikut :

1. *Preprocessing* (Praproses)

Tahap *preprocessing* dilakukan seleksi data, transformasi data dan pembersihan data jika dibutuhkan. Semua entitas data yang berhubungan dengan obyek dan tujuan penelitian harus disiapkan agar siap dimodelkan melalui tahap data mining.

2. *Data Mining*

Tahap data mining pada penelitian ini berupa mengimplementasikan algoritma clustering DBSCAN (*Density-Based Spatial Clustering of Application with Noise*). Tahap ini adalah penggunaan algoritma untuk proses data mining berupa clustering, pengelompokan daerah di karawang berdasarkan data karakteristiknya. Proses data mining pada penelitian ini dilakukan dengan *tools* R Studio, yaitu dengan menggunakan *packages* dbscan.

3. Evaluasi Cluster

Tahap evaluasi cluster dilakukan dengan menggunakan *Average Silhouette width*. Tahap ini ditujukan untuk mengevaluasi hasil pemodelan clustering agar sesuai dengan tujuan dan menjawab rumusan masalah penelitian. Untuk menginterpretasi hasil validasi cluster menggunakan *Average Silhouette width*[8], dapat dilihat rentang nilai dan interpretasi pada tabel I.

Tabel I Rentang nilai *Silhouette* dan interpretasi

Rentang nilai	Interpretasi
0.71-1.0	Struktur yang kuat
0.51-0.70	Struktur yang beralasan
0.26-0.50	Struktur yang lemah
< 0.25	Tidak ditemukan struktur yang substansial

III. HASIL

Hasil dari setiap tahapan penelitian yang sudah dilakukan adalah sebagai berikut :

1. Preprocessing (Praproses)

Tahap *preprocessing* dilakukan seleksi data, transformasi data dan pembersihan data. Seleksi data dilakukan dengan cara memilih 12 atribut/variabel dari dataset yang sudah didapatkan dari Dinas Pertanian Kabupaten Karawang, yaitu kecamatan, luas tanam, produksi, luas panen, luas baku sawah, organisme pengganggu tanaman (penggerek batang, tikus, wereng, batang cokelat, blasit). Transformasi data dilakukan untuk pengembangan dataset sesuai kebutuhan, yaitu penambahan data latitude (garis bujur) dan longitude (garis lintang) untuk menunjukkan lokasi titik kecamatan yang sebelumnya belum ada. Atribut/variabel dataset latitude dan longitude dari semua kecamatan ditambahkan datanya dari [http:// www.mapcoordinates.net/en](http://www.mapcoordinates.net/en).

Selain penambahan terkait atribut kecamatan berupa latitude dan longitude, dilakukan pula penanganan konversi data pada semua atribut dengan cara menyederhanakan satuan, pada variabel luas tanam, produksi, luas panen, luas baku sawah dan luas sawah data dibagi 1000 agar data lebih mudah diolah dan diproses. Hasil transformasi Tabel II dapat dilihat pada Tabel III.

Tabel II. Data Asli Sebelum Dilakukan Transformasi Data

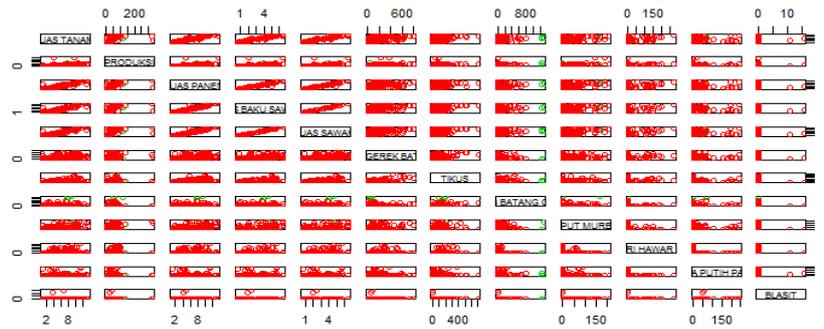
Luas Tanam	Produksi	LuasPanen	LuasBaku Sawah
7600	45564	7600	3814
9862	68416	9862	4931
1507	10797	1507	852
7892	56926	7892	3946
1282	8474	1282	641

Tabel III. Data Asli Setelah Dilakukan Transformasi Data

Luas Tanam	Produksi	LuasPanen	LuasBaku Sawah
7.600	45.564	7.600	3.814
9.862	68.416	9.862	4.931
1.507	10.797	1.507	0.852
7.892	56.926	7.892	3.946
1.282	8.474	1.282	0.641

Pembersihan data (*cleaning data*) dilakukan apabila ditemukan data yang kosong (*missing value*), menghapus data yang tidak diperlukan atau menyeragamkan data yang dianggap sama tetapi ditemukan tidak konsisten pada dataset. Dataset yang digunakan pada penelitian ini tidak ditemukan adanya hal-hal yang menyebabkan pembersihan data perlu dilakukan, dataset lengkap dan data yang dibutuhkan sudah cukup tersedia untuk diolah pada proses data mining.

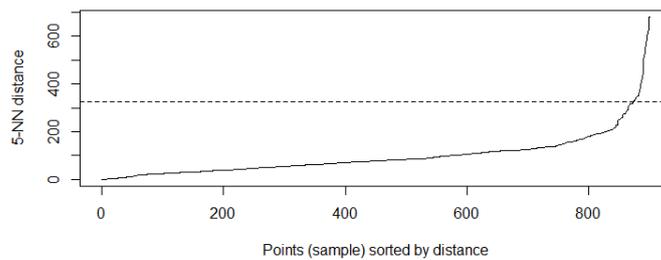
Sebelum dilakukan proses data mining, untuk mengetahui matriks dari scatterplot data dari semua atribut/variabel yang digunakan dapat dilakukan dengan fungsi *pairs()*. Matriks scatterplot dapat disesuaikan dengan pengaturan fungsi panel. Fungsi panel *off-diagonal* dilewatkan kolom yang sesuai dari x sebagai x dan y. Fungsi panel diagonal (jika ada) dilewatkan satu kolom, dan fungsi *text.panel* dilewatkan satu (x, y) lokasi dan nama kolom. Gambar I menunjukkan matriks scatterplot dataset, dimana menunjukkan besar hubungan atau korelasi antar variabel.



Gambar I. Matriks scatterplot dataset

2. Data Mining

Tahap data mining adalah implementasi algoritma clustering DBSCAN (*Density-Based Spatial Clustering of Application with Noise*) pada R studio menggunakan bahasa pemrograman R. Pada tahap ini dilakukan beberapa skenario eksperimen clustering, seperti eksperimen yang dilakukan oleh M.T Furqon dan Lailil M. [1] berdasarkan kombinasi masukan dua parameter yaitu eps dan minPts. Sebelum menentukan berapa eps yang akan digunakan untuk eksperimen DBSCAN, terlebih dahulu menghitung k-Nearest Neighbor Distance (kNNdist) yaitu penghitungan cepat jarak tetangga k-terdekat dalam matriks poin. Plot kNNdist dapat digunakan untuk membantu menemukan nilai eps yang tepat untuk DBSCAN. Pada plot kNNdist akan dicari *knee* karena menunjukkan nilai eps optimal. Berikut ini adalah gambar plot k-NN distance dengan k=5 untuk dataset yang digunakan pada penelitian ini.



Gambar II. Plot k-NN distance

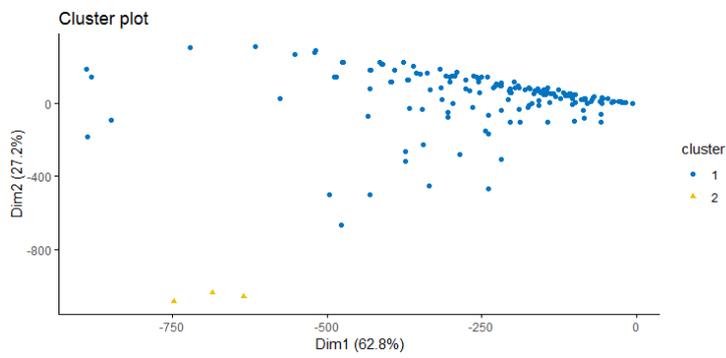
Dari Gambar II diperkirakan *knee* terletak di sekitar angka 300-400, hal ini menunjukkan potensi titik optimal di antara titik itu dan uji coba minPts 1 sampai dengan 5, sehingga eksperimen ditentukan beberapa eps dan minPts. Hasil eksperimen yang menerapkan beberapa skenario eksperimen dapat dilihat pada tabel V.

Tabel IV. Hasil Eksperimen

No	Eps	minPts	Jumlah cluster	Jumlah Noise	Jumlah Tercluster
1	300	1	3	0	180
2	300	2	3	0	180
3	300	3	3	0	180
4	300	4	1	7	173
5	300	5	1	7	173
6	325	1	2	0	180
7	325	2	2	0	180
8	325	3	2	0	180
9	325	4	2	4	176
10	325	5	1	6	174
11	350	1	2	0	180
12	350	2	2	0	180
13	350	3	2	0	180
14	350	4	1	3	177
15	350	5	1	4	176

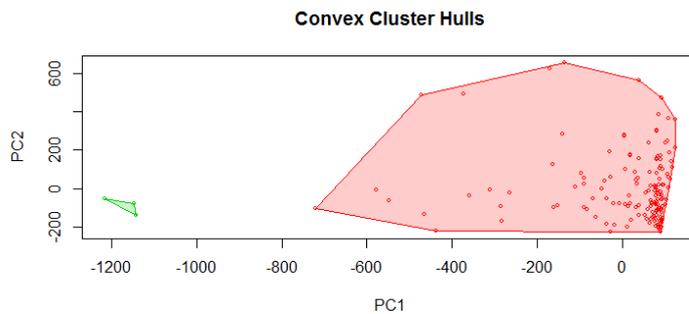
16	375	1	2	0	180
17	375	2	2	0	180
18	375	3	2	0	180
19	375	4	1	3	177
20	375	5	1	4	176
21	400	1	2	0	180
22	400	2	2	0	180
23	400	3	2	0	180
24	400	4	1	3	177
25	400	5	1	4	176

Gambar III merupakan contoh gambar visualisasi hasil clustering pada epsilon = 350 dengan minPts 1 dengan menggunakan fungsi *cluster plot* dan *convex cluster hulls*. Kedua gambar plot di bawah ini menunjukkan plot persebaran dataset yang dapat di-cluster pada akhirnya menjadi dua cluster. Gambar III menunjukkan cluster plot dan Gambar IV menunjukkan hasil cluster menjadi dua.



Gambar III. Cluster plot

Pada gambar III cluster plot menunjukkan sebaran data dikelompokkan menjadi 2 cluster, cluster 1 berwarna merah dan berbentuk bulat sedangkan cluster 2 digambarkan berbentuk segitiga berwarna hijau. Pada Gambar IV menunjukkan area cluster 1 berwarna merah dan area cluster kedua berwarna hijau.



Gambar IV. Gambar convex cluster hulls

IV. PEMBAHASAN

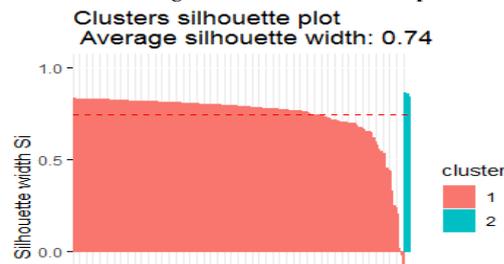
Tahap evaluasi cluster dilakukan dengan menggunakan *Average Silhouette width*. Skor nilai *Average Silhouette width* antara 0 sampai dengan 1, lebih besar dan mendekati 1 menunjukkan kualitas cluster yang lebih baik, yaitu memiliki jarak inter-cluster (jarak antar satu cluster dengan cluster lainnya) tinggi dan jarak intracluster (jarak antar obyek dengan obyek lainnya di dalam satu cluster yang sama) rendah. Hal tersebut menggambarkan bahwa setiap cluster terpisah cukup jauh dengan cluster yang lain sedang setiap anggota cluster memiliki kerapatan yang tinggi dengan obyek-obyek lainnya yang berada dalam satu cluster yang sama. Hasil validasi clustering DBSCAN (*Density-Based Spatial Clustering of Application with Noise*) yang telah dilakukan dapat dilihat pada tabel VI.

Tabel V. Validasi Clustering

No	Epsilon	MinPts	Average silhouette width
1	300	1	0,6

2	300	2	0,6
3	300	3	0,6
4	300	4	0,67
5	300	5	0,67
6	325	1	0,74
7	325	2	0,74
8	325	3	0,74
9	325	4	0,55
10	325	5	0,68
11	350	1	0,74
12	350	2	0,74
13	350	3	0,74
14	350	4	0,74
15	350	5	0,72
16	375	1	0,74
17	375	2	0,74
18	375	3	0,74
19	375	4	0,74
20	375	5	0,72
21	400	1	0,74
22	400	2	0,74
23	400	3	0,74
24	400	4	0,74
25	400	5	0,72

Salah satu hasil cluster yang optimal pada epsilon 325 dengan minPts 2, terdiri dari dua cluster. Cluster 1 terdiri dari 177 item dan cluster 2 hanya 3 item, nilai *average silhouette width* dapat dilihat pada gambar V.



Gambar V. Silhoutte Plot

Hasil cluster dapat digunakan sebagai bahan pemetaan dan bahan analisis rencana penanganan untuk pertanian padi berdasarkan karakteristiknya. Hasil dua cluster pada penelitian ini menunjukkan adanya karakteristik yang berbeda antar cluster, analisis identifikasinya adalah sebagai berikut :

Tabel VI. Identifikasi Karakteristik Hasil Cluster

Cluster	Anggota	Karakteristik
Cluster 1	Kecamatan Cilamaya Kulon, Jatisari, dan Telagasari	- Curah hujan rata-rata tinggi - Jumlah produksi padi tergolong kategori tinggi, karena luas lahan tanam yang luas di tiga kecamatan tersebut.

		- Jumlah serangan hama wereng batang pada lahan sawah tinggi (rata-rata 1269) di tiga kecamatan ini, sedangkan serangan hama penggerek batang (rata-rata 50) .
Cluster 2	Kecamatan Banyusari, Batujaya, Ciampel, Cibuyaya, Cikampek, Cilamaya Wetan, Cilebar, Jayakarta, Karawang Barat, Karawang Timur, Klari, Kota Baru, Kutawaluya, Lemahabang, Majalaya, Pakisjaya, Pangkalan, Pedes, Purwasari, Rawamerta, Rengasdengklok, Tegalwaru, Telukjambe Barat, Telukjambe Timur, Tempuran, Tirtajaya, dan Tirtamulya	- Curah hujan rata-rata lebih rendah - Jumlah produksi padi tergolong kategori sedang, karena lahan sawah ada yang dialihfungsikan menjadi pemukiman/perumahan. - Jumlah hama wereng batang pada lahan sawah lebih rendah (rata-rata 350), sedangkan serangan hama penggerek batang di lahan sawah lebih tinggi (rata-rata 190)

V. SIMPULAN DAN SARAN

Adapun kesimpulan dari penelitian ini adalah sebagai berikut :

1. Penerapan teknik clustering DBSCAN (*Density-Based Spatial Clustering of Application with Noise*) pada lahan pertanian padi di kabupaten Karawang menghasilkan dua cluster yang memiliki perbedaan karakteristik. Hasil cluster menunjukkan perbedaan mengenai tingkat curah hujan, luas lahan yang mempengaruhi jumlah produksi, dan jumlah serangan hama serta jenis hama yang menyerang lahan pertanian.
2. Perbandingan evaluasi performa dari teknik clustering DBSCAN (*Density-Based Spatial Clustering of Application with Noise*) dilakukan dengan memperhatikan nilai dari *average silhouette width*. Besarnya tingkat skor silhouette menunjukkan kualitas cluster yang terbentuk. Hasil eksperimen pada penelitian ini menunjukkan skor hasil tertinggi 0,74 dengan menghasilkan dua cluster. Hal ini menunjukkan cluster yang terbentuk memiliki struktur yang kuat.

Ucapan Terima Kasih

Ucapan terima kasih disampaikan kepada Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Singaperbangsa Karawang yang telah mendukung penuh dan memberikan dana pendukung untuk menyelesaikan penelitian ini.

REFERENSI

- [1] L. M. Furqon, Muhammad Tanzil, "Clustering The Potential Risk Of Tsunami Using Density-Based Spatial Clustering Of Applic", *J. Environ. Eng. Sustain. Technol.*, vol. Vol. 03 No, no. August, pp. 1–8, 2016.
- [2] A. Yuwono, Y. Oslan, S. Kom, and D. D. Dwijono, "Implementasi Metode Density Based Spatial Clustering Of Applications With Noise Noise Border Core Eps = 1cm MinPts = 5."
- [3] N. Made, A. Santika, I. K. Gede, D. Putra, and I. M. Sukarsa, "Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan," vol. 6, no. 3, pp. 185–191, 2015.
- [4] I. U. Dede Nofrianda Utama, Fhira Nhita, "Implementasi Density Based Clustering menggunakan Graphics Processing Unit (GPU)," *e-Proceeding Eng.*, vol. 2, no. 3, pp. 7905–7912, 2015.
- [5] P. Silitonga, "Analisis Pola Penyebaran Penyakit Pasien Pengguna Badan Penyelenggara Jaminan Sosial (Bpjs) Kesehatan Dengan Menggunakan Metode Dbscan Clustering," *J. TIMES*, vol. V, no. 1, pp. 36–39, 2016.
- [6] R. Herteno, "Visualisasi Secara Spasial Cluster Kerusakan Sarana dan Prasarana Sekolah," vol. 8, no. 2, pp. 61–68, 2016.
- [7] J. Zhang, "Web Geospatial Visualisation For Clustering Analysis Of Epidemiological Data," no. December, 2014.
- [8] P. J. Rousseeuw. *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Computational and Applied Mathematics*, 20:53–65, 19