

## Application of K-Means Clustering to Monitoring the level of Fertilizer Usage in Rice Fields in Ngawi Regency

Iwan Ady Prabowo<sup>(1)</sup>; Hendro Wijayanto<sup>(2)</sup>; Sujud Aji Wantoro<sup>(3)</sup>

<sup>1)2)3)</sup> Program Studi Informatika, STMIK Sinar Nusantara

<sup>1)</sup>iwanaadyp@sinus.ac.id; <sup>2)</sup>hendro@sinus.ac.id; <sup>3)</sup>sujudaw@gmail.com;

### ABSTRACT

*Fertilizer is a substance given to plants that could be food for plants. Fertilizer itself divided into two types that is inorganic fertilizers and organic fertilizer. Fertilizer is very important to the needs of farmers, especially farmers. Therefore fertilizer needs should be able to in monitoring in order not to occur lack. The purpose of this study is a grouping of the use of fertilizer in every town in the districts of Ngawi with k-Means algorithm. The formulation of the problem to overcome is how to apply the method of k-means algorithm to categorize fertilizer needs in every town in the Ngawi. K-means is clustering algorithm where one object can be in group based on data similarities. Clustering results K-means Clustering is seen in the first group high (Cluster 1) contains 6, In the second group medium (Cluster 2) contains 2 sub-districts, while in the third group low (cluster 3rd) contains 11 sub-districts.*

**Keywords:** Fertilizer, K-Means, Clustering, Silhouette Coefficient.

### I. PENDAHULUAN

Kabupaten Ngawi terletak di wilayah barat provinsi Jawa Timur yang berbatasan langsung dengan provinsi Jawa Tengah. Wilayah Kabupaten Ngawi adalah 1.298,58 km<sup>2</sup>, di mana sekitar 40% atau sekitar 506,6 km<sup>2</sup> adalah tanah sawah[1]. Tanah berfungsi sebagai sumber nutrisi untuk pertumbuhan tanaman, tempat untuk tumbuh dan menahan akar, dan reservoir untuk air. Beberapa variabel mempengaruhi pertumbuhan tanaman, termasuk ketersediaan nutrisi, baik makro maupun mikro[2][3]. Pupuk adalah salah satu variabel penentu dalam bisnis pertanian[4], ketersediaan pupuk yang cukup dapat membuat tanaman subur dengan memastikan kualitas nutrisi di tanah, sehingga meningkatkan produksi dan kualitas panen secara berkelanjutan[5] [6]. Penelitian ini untuk mengetahui Penerapan k-means clustering untuk monitoring tingkat penggunaan pupuk pada lahan sawah di Kabupaten Ngawi.

Clustering adalah alat yang berguna dalam ilmu data [7]. Metode K-Means adalah salah satu metode dalam fungsi clustering atau pengelompokan berdasarkan kemiripan objek yang diteliti [8].

### II. TINJAUAN PUSTAKA

#### 2.1 Sawah

Sawah adalah lahan yang mencakup hutan, perkebunan, pertanian area, dan yang lainnya. Sawah juga tergantung pada sifat tanah aslinya, karakter sawah sangat bervariasi tergantung pada sifat tanah aslinya[9][10].

#### 2.2 Pupuk

Pupuk adalah satu nutrisi paling penting untuk tanaman. Tanaman membutuhkan nutrisi dalam bentuk mineral dan air setiap hari untuk pertumbuhan, perkembangan, and reproduksi. Akar, batang, dan daun menyerap nutrisi yang dibutuhkan tanaman. Nutrisi ini membuat berbagai tugas yang baik, menjadikannya salah satu komponen terpenting dalam meningkatkan hasil pertanian[4].

#### 2.3 Metode K-Means Clustering

Algoritma K-Means adalah salah satu algoritma pengelompokan data non-hierarkis yang digunakan untuk membagi data menjadi satu atau lebih cluster. Pendekatan ini membagi data menjadi beberapa kelompok, yang mana data dengan sifat serupa dikelompokkan bersama [7][8]. Pendekatan K-Means terdiri dari pengelompokan partisi, yang membagi data menjadi k bagian yang berbeda. Algoritma ini terkenal karena kemudahan penggunaan dan kemampuannya untuk dengan cepat mengklasifikasikan sejumlah besar data. Setiap bagian data harus ditetapkan ke satu cluster, dengan data apa pun milik kluster tersebut diizinkan untuk ditransfer ke kluster lain pada tahap selanjutnya. Pembatasan masalah hanya pada masalah penentuan clustering tanah yang diterapkan untuk mempermudah dalam implementasi sistem[11].

Tahapan dari K-Means Clustering adalah sebagai berikut[8]:

1. Tentukan jumlah cluster.
2. Mendistribusikan data ke dalam kelompok secara acak.

3. Hitung sentroid rata-rata data di setiap kluster.
4. Setiap titik data harus ditetapkan ke sentroid rata-rata yang paling dekat dengannya.
5. Jika masih ada kluster pergeseran data atau jika nilai sentroid berubah, kembali ke langkah 3.

#### 2.4 Penelitian Terkait

Penelitian yang dilakukan oleh Riski Lunika Parmawati dkk[8] dengan judul Clustering Potensi Susu Sapi Perah Di Kabupaten Boyolali Menggunakan Algoritma K-Means membahas tentang mengelompokkan data produksi susu, betina produksi dan pemilik menghasilkan 3 kelompok dengan 13 daerah yang berpotensi susu sapi perah banyak (cluster1), 28 daerah yang berpotensi susu sapi perah sedang (cluster2), dan 28 daerah yang berpotensi susu sapi perah sedikit (cluster3).

Penelitian yang dilakukan oleh Rahmatina Hidayati dkk [12] dengan judul Analisis Silhouette Coefficient pada Perhitungan Jarak K-Means Clustering. Masalah yang dibahas tentang perbandingan jarak K-Means clustering pada data yang memiliki atribut berbeda menggunakan Silhouette coefficient.

### III. METODE PENELITIAN

Penelitian ini adalah studi eksperimental dengan langkah-langkah penelitian seperti analisis, desain, pengkodean, pengujian, dan implementasi.

Tahap analisis sistem melakukan pengamatan langsung terhadap peristiwa yang ditemukan di lapangan seperti luas sawah, jenis irigasi sawah dan lahan yang bukan sawah berdasarkan penggunaannya, pemupukan (jenis pupuk, jumlah pupuk, metode pemupukan, waktu pemupukan), dan kemudian kejadian ini dicatat dan didokumentasikan sebagai data primer. Dari titik ke depan, kegiatan penetapan klasifikasi data pengelompokan penggunaan pupuk dan diambil dari kriteria jenis pupuk, termasuk pupuk organik dan pupuk anorganik, masing-masing kecamatan di Kabupaten Ngawi akan dilakukan.

Use Case Diagram, Kelas Diagram, dan Pemodelan dengan K-Means Clustering merupakan bagian dari *Unified Modelling Language (UML)* yang digunakan untuk desain sistem. Langkah pengkodean ini berkaitan dengan perangkat lunak dan perangkat keras yang digunakan dalam penelitian. Tahap pengujian dan hasil perhitungan dibandingkan untuk menentukan kebenaran K-Means Clustering. Hasil selanjutnya dapat

dimanfaatkan untuk membantu Dinas Pertanian Kabupaten Ngawi dalam memperoleh informasi terkait tingkat kebutuhan pupuk setiap kecamatan di Kabupaten Ngawi.

### IV. HASIL DAN PEMBAHASAN

#### 4.1 Pemodelan sistem dengan Metode K-Means Clustering

Kabupaten Ngawi merupakan salah satu daerah penghasil padi yang produktif. Oleh karena itu harus ditentukan penggunaan pupuk agar tidak ada daerah yang kekurangan pupuk, baik pupuk organik dan anorganik. Metode K-Means digunakan dengan parameter banyaknya kebutuhan pupuk di setiap Kecamatan dengan tujuan untuk mengelompokkan penggunaan pupuk. Dan dibagi menjadi 3 level yaitu, level 1 tinggi, level 2 sedang dan level 3 rendah.

1. Tentukan jumlah cluster.
2. Mendistribusikan data ke dalam kelompok secara acak.
3. Hitung sentroid rata-rata data di setiap kluster.
4. Setiap titik data harus ditetapkan ke sentroid rata-rata yang paling dekat dengannya.
5. Jika masih ada kluster pergeseran data atau jika nilai sentroid berubah, kembali ke langkah 3.

##### 4.1.1 Tentukan jumlah cluster

Penentuan cluster diambil dari tiga jenis kebutuhan kecamatan yaitu, Tinggi (C1), Sedang (C2), dan Rendah (C3). Jumlah data diambil dari data pemupukan per Kecamatan di Kabupaten Ngawi merujuk pada Tabel 1. sehingga diperoleh data sejumlah 19 data. Jumlah atribut diambil dari kriteria jenis pupuk yang akan mempengaruhi sirkulasi Clustering. Jadi kriteria dijadikan syarat untuk mengetahui tingkat kebutuhan pupuk per Kecamatan di Kabupaten Ngawi. Percobaan dilakukan dengan menggunakan parameter-parameter berikut: (a) Jumlah cluster = 3, (b) Jumlah data = 19, Jumlah atribut : 2.

Berikut merupakan data yang diperoleh dari Dinas Pertanian.

**Tabel 1. Kebutuhan Pupuk di Kabupaten Ngawi**

No	Kecamatan	Anorganik	Organik
1	1	66.693	402.047
2	2	31.136	182.848
3	3	49.230	259.581
4	4	50.577	289.494
5	5	51.083	281.392

No	Kecamatan	Anorganik	Organik
6	6	65.597	464.498
7	7	72.552	387.168
8	8	162.619	964.001
9	9	113.609	706.419
10	10	61.330	329.557
11	11	48.902	290.954
12	12	59.130	311.258
13	13	64.723	250.332
14	14	180.330	350.302
15	15	47.230	300.215
16	16	110.223	270.450
17	17	50.143	604.273
18	18	46.301	320.320
19	19	34.210	34.210
...	...	...	...

4.1.2 Pusat Cluster Awal

Pusat awal cluster dapat ditentukan bebas dengan mengambil tiga data yang ada. Penentuan pusat awal cluster dengan sebagai berikut:

- pusat cluster ke-1 di ambil data ke-14
- pusat cluster ke-2 di ambil data ke-8
- pusat cluster ke-3 di ambil data ke-2

Tabel 2. Hasil Jarak Antar Cluster Data 1

No	C1	C2	C3
1	124863,5767	570082,5	124863,5767
2	224275,9277	792141,3	0
3	159428,6983	713487,6	78837,47
4	143294,989	683749,3	108403,5
5	146469,7071	691661,3	100542,5
6	161877,6937	508838,4	283750,4
7	113908,7321	583822,2	208475,3
8	613954,5114	0	792141,3
9	362313,4134	262203,1	530026,8
10	120794,6813	642478,5	149783,9
11	144206,4641	682586,1	109556,1
12	127333,7109	660895,9	131426
13	152836,4464	720352	75380,22
14	0	613954,5	224275,9
15	142212,2272	673740,7	118465,3
16	106260,6858	695527,4	118020,6
17	285394,3304	376902	421853,4
18	137341,5202	654106,3	138305,9
19	177073,021	529486,6	267490,7
...	...	...	...

4.1.3 Perhitungan Jarak Pusat Cluster

Untuk mengukur jarak antara tiap cluster dengan pusat cluster digunakan *Euclidian Distance* seperti pada persamaan (1). Kemudian akan didapatkan matriks jarak yaitu C1, C2 dan C3 diuraikan pada Tabel 2. Perhitungan sebagai berikut:

- Jarak antara data Kecamatan pertama dengan pusat cluster pertama

$$C1 = (66.693-180.330)^2 + (402.047-350.302)^2 = 15.590.912.786,31678$$

$$15.590.912.786,31678 = 124.863,5767$$

- Jarak antara data Kecamatan pertama dengan pusat cluster kedua

$$C2 = (66.693-162.619)^2 + (402.047-964.001)^2 = 324.994.056.806,25$$

$$324.994.056.806,25 = 570.082,5$$

- Jarak antara data Kecamatan pertama dengan pusat cluster ketiga

$$C3 = (66.693-31.136)^2 + (402.047-182.848)^2 = 49.312.508.921,64$$

$$49.312.508.921,64 = 222.064,2$$

4.1.4 Pengelompokan Data

Jarak antara hasil perhitungan dan pusat cluster akan dibandingkan dan jarak terdekat antara data dan pusat cluster akan dipilih; jarak ini menunjukkan bahwa data berada dalam satu kelompok dengan pusat kluster terdekat.

Berikut ini akan ditampilkan data matriks pengelompokan grup, simbol “ \* “ berarti data tersebut berada dalam grup (kelompok data). Tabel 3 menunjukkan, Data 1 memiliki jarak terpendek di cluster C1, data 2 memiliki jarak terpendek di cluster C3, dan seterusnya.

Tabel 3. Pengelompokan Cluster Data 1

No	C1	C2	C3
1	*		
2			*
3			*
4			*
5			*
6	*		
7	*		
8		*	
9		*	
10	*		
11			*
12	*		
13			*
14	*		
15			*

No	C1	C2	C3
16	*		
17	*		
18	*		
19	*		
...	...	...	...

4.1.5 Penentuan Pusat Cluster Baru

Setelah diketahui anggota masing-masing kluster maka pusat kluster baru dihitung berdasarkan data anggota masing-masing kluster sesuai dengan rumus pusat anggota kluster. Sehingga mendapatkan perhitungan berikut:

Berikut uraian perhitungan data untuk perhitungan kelompok data iterasi ke-2 :

- Untuk cluster pertama, ada 10 data yaitu data ke 1, 6, 7, 10, 12, 14, 16, 17, 18, 19 sehingga:

$$C1.1 = (66.693+65.597+72.552+61.330+59.130 +180.330+110.223+50.143+46.301+34.210) / 10 = 74.650,9$$

$$C1.2 = (402.047+464.498+387.168+329.557+3 11.258+350.302+270.450+604.273+320.320+450.321) / 10 = 389.019,4$$

- Untuk cluster kedua, ada 2 data yaitu data ke 8,9 sehingga:

$$C2.1 = (162.619+113.609) / 2 = 138.114$$

$$C2.2 = (964.001+706.419) / 2 = 835.210$$

- Untuk cluster ketiga, ada 7 data yaitu data ke 2, 3, 4, 5, 11, 13, 15 sehingga :

$$C3.1 = (31.136+49.230+50.577+51.083+48.902 +64.723+47.230) / 7 = 48.983$$

$$C3.2 = (182.848+259.581+289.494+281.392+2 90.954+250.332+300.215) / 7 = 264.973,7$$

Pada Tabel 4, pembentukan data kluster baru berasal dari pembentukan iterasi pusat kluster baru 2.

Tabel 4. Pembentukan Cluster Baru

Cluster Baru		
C1	C2	C3
74.650,9	138.144	48.983
389.019,4	835.210	264.873,7

4.1.6 Ulangi Langkah Kedua Hingga Posisi Data Tidak Mengalami Perubahan

Matriks jarak yaitu C1, C2 dan C3 pada iterasi 2 diuraikan pada Tabel 5.

Tabel 5. Pengelompokan Data (Kelompok Cluster Iterasi 2, Iterasi 3, Iterasi 4, Iterasi 5)

No	Iterasi 2			Iterasi 3			Iterasi 4			Iterasi 5		
	C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
1	*			*			*			*		
2			*			*			*			*
3			*			*			*			*
4			*			*			*			*
5			*			*			*			*
6	*			*			*			*		
7	*			*			*			*		
8		*			*			*			*	
9		*			*			*			*	
10	*					*			*			*
11			*			*			*			*
12			*			*			*			*
13			*			*			*			*
14	*			*			*			*		
15			*			*			*			*
16			*			*			*			*
17	*			*			*			*		
18			*			*			*			*
19	*			*			*			*		
...	...	...	...	...	...	...	...	...	...	...	...	...

Pada Tabel 5, Data tidak berubah sebagai hasil dari iterasi hingga yang kelima. Akibatnya, iterasi berakhir pada langkah iterasi kelima. Hasilnya, tingkat kebutuhan pupuk suatu kecamatan dapat ditentukan, mulai dari rendah hingga sedang dan hingga tinggi.

4.2 Pengujian dan Implementasi Sistem

Metode Silhouette Coefficient, metode pengukuran yang menggabungkan kohesi dan pemisahan, digunakan untuk pengujian metode. Proses tiga tahap digunakan untuk menghitung Silhouette Coefficient untuk satu titik tunggal[12].

1. Pada objek ke i, dihitung jarak rata-rata terhadap semua objek yang lain dalam kluster, kemudian disebut dengan a<sub>1</sub>.
2. A Pada objek ke i, dihitung jarak rata-rata terhadap semua objek lain di kluster lain, yang kemudian disebut dengan b<sub>1</sub>
3. Koefisien siluet untuk suatu titik didapatkan dengan:

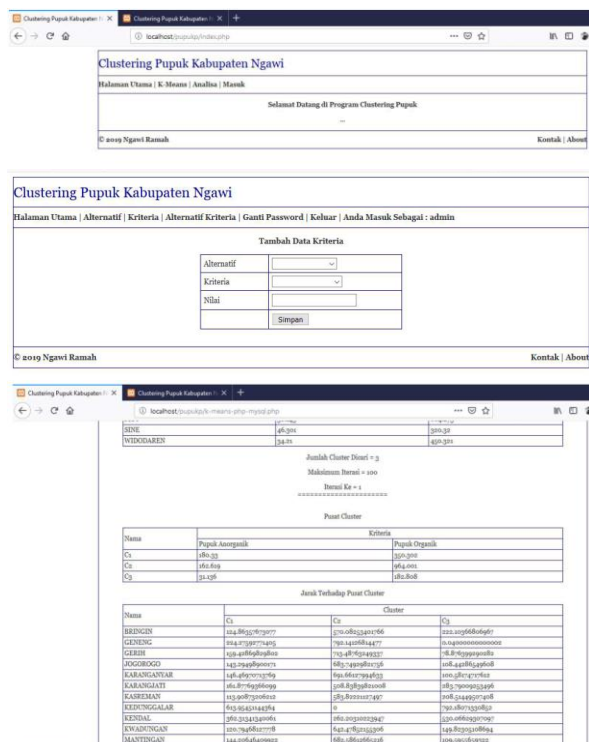
$$s=1 - a/b \text{ jika } a < b \text{ } s= b/a -1 \text{ jika } a \geq b$$

Nilai koefisien siluet bervariasi antara -1 dan 1. Nilai negatif tidak diinginkan sebab a<sub>1</sub> adalah rata-rata jarak dari titik ke kluster. Nilai koefisien rata-rata setiap titik yang paling baik adalah yang semakin mendekati angka 1.

Tabel 6. Koefisien Silhouette

Kecamatan	A(i)	B(i)	S(i)
1	201189,1289	158537,8389	0,269029
2	342164,6491	241230,3559	0,418415
3	263475,9588	241300,4768	0,0919
4	239797,5011	229178,2432	0,046336
5	243307,081	237976,7993	0,022398
6	195582,1222	170722,2332	0,145616
7	205615,7056	149392,3018	0,376347
8	450140,1427	319038,5916	0,410927
9	275483,7308	144382,1797	0,908018
10	225158,6707	175931,1748	0,279811
11	239747,3666	226977,3514	0,056261
12	231298,9338	200497,2557	0,153626
13	271365,1254	239978,1648	0,130791
14	255644,5646	117690,866	0,17217
15	236841,323	215192,8234	0,1006
16	265283,5726	210467,4364	0,260449
17	244826,7862	81222,10994	0,106701
18	230290,7778	189301,4416	0,216529
19	204825,4315	160193,4945	0,278613
...	...	...	...

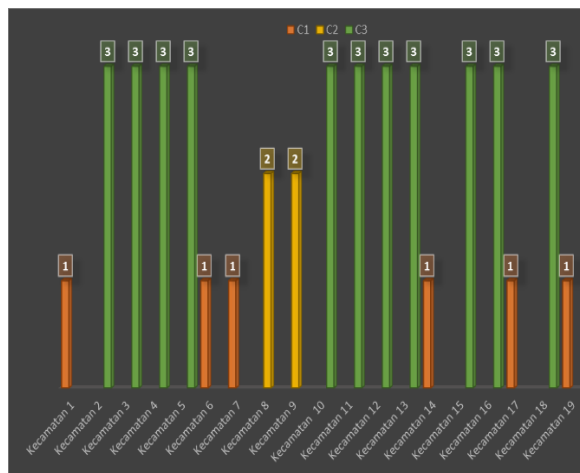
Implementasi sistem merupakan proses pembuatan sistem berupa Aplikasi Clustering penggunaan Pupuk Kabupaten Ngawi. Adapun tampilannya sebagai Gambar 1.



Gambar 1. Aplikasi Clustering penggunaan Pupuk Kabupaten Ngawi

Hasil Clustering *K-means Clustering* terlihat pada Kelompok pertama tinggi (Cluster ke-1)

berisi 6 Kecamatan yaitu : Kecamatan 1, Kecamatan 6, Kecamatan 7, Kecamatan 14, Kecamatan 17, Kecamatan 19. Pada Kelompok kedua sedang (Cluster ke-2) berisi 2 Kecamatan yaitu Kecamatan 8 dan Kecamatan 9. Sedangkan pada kelompok ketiga rendah (Cluster ke-3) berisi 11 Kecamatan yaitu: Kecamatan 2, Kecamatan 3, Kecamatan 4, Kecamatan 5, Kecamatan 10, Kecamatan 11, Kecamatan 12, Kecamatan 13, Kecamatan 15, Kecamatan 16, Kecamatan 18. Hasil Clustering terlihat pada Gambar 8.



Gambar 8. Grafik hasil Cluster masing-masing Kecamatan dengan *K-means Clustering*

V. PENUTUP

Berdasarkan hasil pembahasan maka kesimpulannya sebagai berikut:

Telah terbangunnya aplikasi clustering dengan menggunakan metode k-means untuk menentukan beberapa kelompok cluster pupuk di Kabupaten Ngawi yaitu C1 (Rendah), C2(Sedang), dan C3(Tinggi). Penelitian ini masih mempunyai kekurangan, masih dapat dilakukan untuk mengembangkan aplikasi ini agar menjadi lebih baik, saran dalam penelitian ini sistem tidak mendeteksi perubahan kondisi pupuk, dan jumlah pupuk yang terus berubah.

DAFTAR PUSTAKA

- [1] S. J. Ariyanto, and A. Permana Hakim, "Kabupaten Ngawi Dalam Angka 2021," Ngawi, Feb. 2021.
- [2] H. D. Foth and B. G. Ellis, "Soil Fertility and Plant Nutrition," *Soil Fertility*, pp. 1–24, 2018, doi: 10.1201/9780203739341-1.
- [3] A. S. Jadeja, D. v Hirpara, L. C. Vekaria, and H. L. Sakarvadia, "Soil Fertility Evaluation," *Soil Fertility and Nutrient Management*, pp. 64–75, 2021, doi: 10.1201/9781003200239-5.

- [4] A. Wihardjaka, "Dukungan Pupuk Organik Untuk Memperbaiki Kualitas Tanah Pada Pengelolaan Padi Sawah Ramah Lingkungan," *Jurnal Pangan*, 2021, [Online]. Available: <http://www.jurnalpangan.com/index.php/pangan/article/view/496>
- [5] I. A. Adi, N. Barunawati, and T. Wardiyati, "Pengaruh kombinasi pupuk NPK dengan jenis pupuk kandang pada pertumbuhan dan hasil kentang (*Solanum tuberosum* L) di dataran medium," *Jurnal Produksi Tanaman*, 2017.
- [6] C. Yolanda Apriscia, N. P. Barunawati dan Karuniawan Wicaksono Jurusan Budidaya Pertanian, and F. Pertanian, "Pengaruh Pemberian Pupuk Kompos Limbah Domestik Terhadap Pertumbuhan Vegetatif Tanaman Tebu (*Saccharum officinarum* L.) Asal Bibit Bud Chip."
- [7] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [8] R. L. Parmawati, I. A. Prabowo, and T. Susyanto, "Clustering Potensi Susu Sapi Perah Di Kabupaten Boyolali Menggunakan Algoritma K-MeansK-MEANS," *Jurnal Teknologi Informasi dan Komunikasi (TIKomsin)*, vol. 7, no. 1, Jul. 2019, doi: 10.30646/tikomsin.v7i1.413.
- [9] S. Hardjowigeno, H. Subagyo, and M. L. Rayes, "Morfologi dan klasifikasi tanah sawah," *Di dalam Tanah Sawah dan Teknol pengelolaannya Pus Penelit Tanah dan Agroklimat Dep Pertan Bogor*, 2004.
- [10] R. P. Chandra, *Identifikasi Konversi Lahan Sawah Di Kecamatan Ciparay, Kabupaten Bandung*. elibrary.unikom.ac.id, 2020. [Online]. Available: <https://elibrary.unikom.ac.id/id/eprint/3884/>
- [11] O. Somantri, S. Wiyono, and D. Dairoh, "Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM)," *Scientific Journal of Informatics*, vol. 3, no. 1, pp. 34–45, 2016.
- [12] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering," *Techno. Com*, 2021, [Online]. Available: <http://publikasi.dinus.ac.id/index.php/technoc/article/view/4556>