

Klasterisasi Wilayah Penghasil Tanaman Lada Menggunakan Algoritma K-Means**Novianti Puspitasari¹, Havaluddin², Fazma Urmila Jannah Helmi Puadi³**

novia.ftik.unmul@gmail.com, havaluddin@unmul.ac.id, fazmaurmila@gmail.com

Universitas Mulawarman

Informasi Artikel

Diterima : 16 Okt 2022

Direview : 30 Nov 2022

Disetujui : 5 Des 2022

Kata KunciLada, *K-Means*, wilayah, *Silhouette Coefficient*, *Sum Square Error***Abstrak**

Wilayah potensial untuk menanam lada semakin berkurang, sehingga jumlah produksi lada menjadi semakin menurun. Hal ini tentunya perlu menjadi perhatian mengingat lada merupakan salah satu komoditas unggulan yang sangat penting untuk menunjang perekonomian. Informasi tentang daerah yang berpotensi sebagai daerah penghasil tanaman lada perlu dilakukan. Penelitian ini bertujuan untuk mendata dan menganalisa wilayah potensial untuk tanaman lada menggunakan pendekatan algoritma cerdas yaitu *K-Means*. Data penelitian berasal dari Dinas Perkebunan Provinsi Kalimantan Timur sebanyak 1200 data dalam rentang waktu tahun 1990 sampai 2019 telah digunakan untuk dianalisis. Lebih lanjut, ketiga metode jarak yaitu *Euclidean Distance*, *Manhattan Distance* dan *Minkowski Distance* digunakan dalam penelitian ini. Dari ketiga metode tersebut dicari nilai akurasi yang tertinggi menggunakan metode *Silhouette Coefficient* (SC). Metode *Sum Square Error* (SSE) dan *R-squared* (R^2) juga digunakan untuk mengukur *cluster* optimal. Hasil percobaan memperlihatkan bahwa metode jarak *Manhattan Distance* memiliki nilai akurasi terbaik. Sedangkan, *cluster* optimal untuk klusterisasi wilayah diperoleh tiga *cluster* yang merupakan *cluster* ideal untuk mengelompokkan wilayah penanam lada dengan SSE sebesar 238.7377116 dan nilai R^2 adalah 0.459398609. Berdasarkan hasil tersebut, diperoleh informasi tentang wilayah yang berpotensi untuk produksi lada menggunakan tiga kategori yaitu kurang berpotensi, cukup berpotensi dan berpotensi baik dengan algoritma *K-Means* dan metode jarak *Manhattan Distance*.

Keywordspepper, *K-Means*, area, *Silhouette Coefficient*, *Sum Square Error***Abstrak**

The potential area for growing pepper is decreasing, so the amount of pepper production is declining. It undoubtedly needs to be a concern considering that pepper is one of the leading commodities that is very important to support the economy. Information about areas that have the potential to produce pepper plants needs to be done. This study aims to record and analyze potential areas for pepper plants using an intelligent algorithm approach, namely K-Means. Research data from the Plantation Office of East Kalimantan Province, 1200 data from 1990 to 2019 have been used for analysis. Furthermore, the three distance methods, namely Euclidean Distance, Manhattan Distance, and Minkowski Distance, are used in this study. Of the three methods, the highest accuracy value is sought using the Silhouette Coefficient (SC) method. Sum Square Error (SSE) and R-squared (R²) methods were also used to measure the optimal cluster. The experimental results show that the Manhattan Distance method has the best accuracy value. Meanwhile, the optimal cluster for area clustering is obtained from three groups, ideal clusters for classifying pepper growing areas with an SSE of 238.7377116 and an R² value of 0.459398609. Based on these results, information was obtained about places with the potential for pepper production using the K-Means algorithm and the Manhattan Distance method using three categories, namely less potential, moderately potential, and good potential.

A. Pendahuluan

Lada merupakan komoditas tradisional yang sudah lama dikenal dan dikembangkan oleh masyarakat, karena tanaman ini merupakan 10 (sepuluh) komoditi unggulan untuk sektor perkebunan. Provinsi Kalimantan Timur merupakan salah satu pengeksport lada nasional karena lahannya yang cocok untuk ditanami. Sebagai contoh wilayah Kabupaten Kutai Kartanegara menanam dan mengolah lada yang dikenal dengan istilah "*Samarinda white-pepper*" karena berasal dari jenis yang berbeda dan bermutu [1]. Namun, saat ini produksi lada mengalami penurunan [2]–[4]. Komoditi lada mengalami penurunan produksi akibat berkurangnya luas tanam produktif dan minimnya pemeliharaan tanaman. Kendati telah dilakukan kegiatan ekstensifikasi atau perluasan lahan, komoditi lada tetap mengalami penurunan luas tanam yang cukup drastis akibat alih fungsi lahan menjadi kelapa sawit, tanaman pangan dan sektor pertambangan [2]. Oleh karena itu, upaya peningkatan produksi lada harus digalakkan kembali untuk mendukung perekonomian provinsi dan nasional melalui sektor perkebunan. Upaya yang dapat dilakukan untuk meningkatkan produksi lada adalah menentukan tingkat potensi suatu wilayah sebagai penghasil tanaman lada. Wilayah-wilayah tersebut nantinya diklusterisasi ke dalam beberapa kategori seperti berpotensi atau tidak berpotensi.

Berbagai algoritma cerdas telah dan terus diimplementasikan dalam berbagai sektor salah satunya adalah algoritma *K-Means*. Algoritma *K-Means* sudah pernah digunakan pada sektor pertanian untuk mengelompokkan wilayah yang potensial dalam menghasilkan hasil pertanian. Sistem yang dibangun menggunakan *K-Means* memiliki validasi sebesar 100%. Berdasarkan hal tersebut, algoritma *K-Means* dapat dijadikan alternatif acuan dalam menentukan potensi suatu wilayah untuk lahan pertanian [5]. Di sektor perkebunan, *K-Means* mampu memberikan informasi terkait daerah potensial untuk menanam tanaman kelapa hibrida disepuluh desa dengan kategori kurang potensial sebanyak 7 desa, sangat berpotensi 2 desa dan 1 desa yang cukup berpotensi [6]. Lebih lanjut, pengelompokan tentang potensial suatu kawasan untuk pertanian karet produktif juga telah menerapkan algoritma *K-Means* dengan hasil penelitian menunjukkan bahwa algoritma ini mampu mengelompokkan kawasan yang baik dan kurang baik untuk produksi karet [7]. Pengelompokan potensi komoditas perkebunan yang menerapkan *K-Means* ke dalam sistem menghasilkan tingkat akurasi sebesar 93,75% untuk metode pengukuran jarak *Euclidean Distance* dan *Sum of Squared Error* (SSE) sebesar 6.25%. Dari hal tersebut terlihat bahwa *K-Means* cukup akurat dalam mengelompokkan komoditas perkebunan, karena nilai SSE tidak melebihi 50% [8]. Di sektor lainnya, *K-Means* juga mampu menentukan status gizi balita [9], mengklusterisasi pola penyebaran penyakit dan penyakit pasien [10], [11], mengelompokkan curah hujan di wilayah Kalimantan Timur, wilayah pendistribusian listrik, daerah rawan kriminalitas dan tingkat penyebaran kasus COVID-19 di Indonesia [12]–[15].

Berdasarkan penerapan *K-Means* dalam berbagai penelitian dan bidang, terlihat bahwa metode *K-Means* memiliki keunggulan yaitu relatif terukur dan memiliki ketelitian yang cukup baik terhadap ukuran data [16]–[18]. Penelitian ini bertujuan untuk mengklusterisasi wilayah penanam lada menggunakan algoritma *K-Means* dengan menggunakan tiga metode pengukuran jarak. Informasi tentang

klusterisasi wilayah penghasil lada diharapkan dapat membantu pihak-pihak terkait seperti Dinas Perkebunan untuk memaksimalkan wilayah-wilayah yang berpotensi meningkatkan produksi tanaman lada.

B. Metode Penelitian

Adapun tahapan penelitian yang harus dilakukan adalah sebagai berikut:

1. Metodologi Penelitian

Metodologi penelitian yang digunakan dalam penelitian ini terdiri dari:

- a. Studi literatur dan Studi Pustaka: mengumpulkan bahan dan informasi yang berkaitan dengan algoritma *clustering K-Means* dari beberapa jurnal dan juga penelitian-penelitian sebelumnya.
- b. Identifikasi Masalah: mengidentifikasi masalah berdasarkan hasil dari informasi yang diperoleh dari jurnal dan penelitian sebelumnya yang berkaitan dengan algoritma *K-Means*. Permasalahan yang diangkat adalah turunnya produksi lada di Provinsi Kalimantan Timur.
- c. Menentukan Variabel dan Sumber Data: menentukan variabel yang akan digunakan dalam penelitian. Variabel produksi dan luas lahan atau luas areal usaha tani lada merupakan variabel yang sesuai dengan masalah yang diangkat dalam penelitian, sedangkan variabel lainnya akan diambil dari hasil analisis variabel ketika telah memperoleh data penelitian.
- d. Observasi Lapangan: mengunjungi instansi Dinas Perkebunan Provinsi Kalimantan Timur untuk memastikan ketersediaan data dan melakukan pengambilan data.
- e. Pengumpulan Data: mengunjungi *website* resmi Dinas Perkebunan Provinsi Kalimantan Timur untuk mendapatkan data laporan statistik produksi dan produktivitas lada serta melakukan wawancara dengan pihak PPID Dinas Perkebunan Provinsi Kalimantan Timur. Wawancara dilakukan dengan mengajukan beberapa pertanyaan mengenai data statistik perkebunan yang telah diperoleh.
- f. Analisis dan Pengolahan Data: mengelola data produksi dan produktivitas lada menggunakan Ms. *Excel* kemudian menyiapkan data tersebut sehingga siap untuk digunakan dalam penelitian. Setelah mengelola data, data tersebut dianalisis menggunakan metode *Crisp-DM*.
- g. Implementasi *K-Means*: melakukan proses *clustering* menggunakan algoritma *K-Means* terhadap data yang telah dianalisis sehingga membentuk *cluster* yang telah ditetapkan.
- h. Penarikan Kesimpulan: menarik kesimpulan berdasarkan hasil dari penelitian yang telah dilakukan.

2. Algoritma *K-Means*

Clustering merupakan proses maupun metode untuk menganalisis data yang melakukan pengelompokan satu set fisik atau objek abstrak ke dalam kelas-kelas objek yang serupa [19], [20]. Lebih lanjut, *clustering* bertujuan untuk mengelompokkan data dengan karakteristik yang mirip dengan area yang sama dan data dengan karakteristik berbeda ke area berbeda [21]. Algoritma *clustering* yang digunakan dalam penelitian ini adalah *K-Means*. *K-Means* sangat populer dan

telah dimanfaatkan diberbagai bidang. Algoritma ini memiliki konsep yang sederhana dengan ketelitian yang cukup baik terhadap ukuran data. K-Means relatif terukur dan efisien untuk pengolahan data dalam jumlah besar tanpa terpengaruh terhadap urutan data. Komputasi yang dimiliki oleh K-Means relatif cepat [16]–[18]. Algoritma ini didasarkan pada penentuan jarak antara *centroid* dan data latih. Penentuan jumlah *cluster centroid* didasarkan dari jumlah *cluster* yang diinginkan. Sementara itu, inisialisasi *centroid* dilakukan secara acak dengan mempertimbangkan data latih yang digunakan. Selanjutnya setiap iterasi akan dihitung jarak setiap data latih dengan setiap *centroid*. Pada waktu yang bersamaan, anggota *cluster* ditunjukkan oleh jarak terkecil dari *centroid*. Kemudian nilai *centroid* baru dihitung berdasarkan nilai rata-rata dari setiap anggota *cluster*. Jika anggota *cluster* tidak berubah maka iterasi dihentikan [22].

3. Perhitungan Jarak

Jumlah *cluster* awal yang ditetapkan dalam penelitian ini sebanyak 3 kategori wilayah yang terdiri dari berpotensi baik (BB), berpotensi sedang (BS) dan kurang berpotensi (BK). Selanjutnya, menentukan *centroid* secara random dan menghitung jarak terdekat dengan *centroid*. Dalam penelitian ini, 3 metode perhitungan jarak telah dieksplorasi. Metode tersebut terdiri dari *Euclidean Distance* dengan Persamaan 1 dan *Manhattan Distance* dengan Persamaan 2 [23]. Lebih lanjut *Minkowski Distance* dengan Persamaan 3 [24].

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

Dimana, d adalah jarak antara x dan y ; x adalah pusat *cluster*; y adalah data pada atribut; i adalah setiap data; n adalah jumlah data; x_i adalah data pada pusat *cluster* i ; dan y_i adalah data pada pusat data ke i .

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

Dimana, d adalah jarak antara x dan y ; x adalah pusat *cluster*; y adalah data pada atribut; i adalah setiap data; n adalah jumlah data; x_i adalah data pada pusat *cluster* i ; dan y_i adalah data pada pusat data ke i .

$$d(x, y) = (\sum_{i=1}^n |x_i - y_i|^p)^{1/p} \quad (3)$$

Dimana, d adalah jarak antara x dan y ; x adalah data pusat *cluster*; y adalah data pada atribut; i adalah setiap data; n adalah jumlah data; x_i adalah data pada pusat *cluster* ke i ; dan y_i adalah data pada pusat data ke i dan p adalah power.

4. Perhitungan Akurasi

Penelitian ini menggunakan 3 metode uji akurasi. Metode pertama yaitu *Silhouette Coefficient* yang merupakan metode untuk melihat kualitas dan kekuatan *cluster* dengan Persamaan 4 [23], [25].

$$S_{(i)} = \frac{b_{(i)} - a_{(i)}}{\max(a_{(i)}, b_{(i)})} \quad (4)$$

Dimana, $a_{(i)}$ adalah rata-rata jarak suatu objek ke- i dengan semua objek lain pada satu *cluster*; $b_{(i)}$ adalah rata-rata jarak dari objek ke- i dengan semua objek pada *cluster* lain; $S_{(i)}$ adalah nilai SC.

Selanjutnya, metode *Sum Square Error* (SSE) sebagai metode yang digunakan untuk mengukur selisih total dari nilai sebenarnya terhadap nilai yang tercapai seperti terlihat pada Persamaan 5 [2], [26], [27].

$$SSE = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (5)$$

Dimana, x_i adalah nilai data tiap *cluster* i ; \bar{x} adalah nilai rata-rata data *cluster*. Metode ketiga adalah metode *R Squared* (R^2) yang merupakan metode untuk mengecek seberapa tinggi mutu klasterisasi dengan perhitungan seperti Persamaan 6 [28], [29].

$$R_{squared} = \frac{SS_b}{SS_t} \quad (6)$$

Dimana SS_b merupakan jumlah kuadrat antar kelompok; dan SS_t merupakan jumlah kuadrat total.

5. Data Penelitian

Pada penelitian ini, data penelitian yang digunakan adalah data sekunder yang diperoleh dari *website* resmi Dinas Perkebunan Provinsi Kalimantan Timur dengan alamat <https://disbun.kaltimprov.go.id> dan data primer dari wawancara dengan pihak terkait. Data yang tersedia sebanyak 1200 data yang berasal dari laporan statistik tahun (1990-2019) seperti terlihat pada Tabel 1.

Tabel 1. Data Penelitian

Tahun : 1990								
No	Provinsi/ Kabupaten/ Kota	Luas Areal (Ha)			Jumlah (Ha)	Produksi (Ton)	Rata-Rata Produksi (Kg/Ha)	Tenaga Kerja Perkebunan (TKP)
		TBM	TM	TT/ TR				
1	Samarinda	244,5	141	76	461,5	83	590	384
2	Balikpapan	201	27	0	228	22	815	220
3	Kutai	3916,5	1929	105	6950	3005,5	770	7922
4	Pasir	248	37	7	292	124	1162	224
5	Berau	345	302	8	655	105	347	427
6	Bulungan	162	27	0	189	12	426	100
7	Jumlah	5117	2463	196	8776	3351	944	9277
...
...
Tahun : 2019								
1	Kutai							
1	Kartanegara	465	3417	331	4213	3514	1028	2186
2	Kutai Timur	254	166	0	420	110	660	687
3	Kutai Barat	66	16	30	112	5	311	92
4	Mahakam Ulu	3	-	-	3	-	-	2
5	Penajam (P.U)	213	1143	-	1356	1006	880	1740
6	Paser	24	62	10	96	13	205	295
7	Berau	1039	1439	91	2569	1065	740	2566
8	Samarinda	12	26	4	42	12	462	57
9	Balikpapan	25	82	1	108	74	902	119
10	Bontang	2	-	-	2	-	-	4

Dari Tabel 1. terlihat bahwa dari 16 nama kabupaten atau kota yang pernah menjadi bagian dari Provinsi Kalimantan Timur pada tahun 1990 s/d 2019 mengalami perubahan karena terdapat daerah yang mengalami pemekaran. Hal ini menyisakan 10 Kabupaten atau Kota yang merupakan daerah potensial untuk

ditanami tanaman lada. Selain itu untuk mengatasi masalah data seperti inkonsistensi dan hilang maka dilakukan serangkaian proses CRISP-DM berupa *cleaning* dan transformasi. Hasil data yang telah dilakukan proses CRISP-DM dapat dilihat pada Tabel 2.

Tabel 2. Transformasi Data

Kabupaten/ Kota	Luas Lahan (Ha)	Produksi (Ton)	Rata-Rata Produksi (Ton/Ha)	Tenaga Kerja Perkebunan (TKP)
K1	229358.69	140154.1	30.902506	142273
K2	92994.1933	29096.1	22.533736	80248
K3	88042.6933	26903.1	16.387766	70191
K4	2924.15644	896.37	0.463658867	2324.57
K5	37042.9	22495.95833	20.71251833	61384.67
K6	12059.9	2279.458333	14.59527833	0
K7	37015.5417	21438.83333	24.7547475	38255.5
K8	7012.83333	2548.75	18.66120833	9702.166667
K9	3655.375	1885.291667	34.11945833	7485.75
K10	86365.6933	26725.6	8.195816	68205

Keterangan:

Kabupaten/Kota terdiri dari 10 yaitu K1=Kutai Kartanegara, K2=Kutai Timur, K3=Kutai Barat; K4=Mahakam Ulu; K5=Penajam Paser Utara; K6=Paser; K7=Berau; K8=Samarinda; K9=Balikpapan; K10=Bontang.

Tabel 2. menampilkan parameter yang menjadi penentu dalam mengklusterisasi wilayah penanam lada. Parameter tersebut adalah luas area tanaman lada (luas lahan), jumlah produksi tanaman lada (produksi), jumlah rata-rata produksi tanaman lada dan jumlah tenaga kerja di perkebunan tanaman lada (TKP). Keempat parameter tersebut didasarkan pada hasil wawancara dengan Dinas Perkebunan. Setelah data wilayah ditransformasi selanjutnya melakukan normalisasi menggunakan metode *z-score* untuk menyamakan rentang nilai menggunakan Persamaan 7 [27].

$$Z_i = \frac{x_i - \bar{x}}{s} \quad (7)$$

Dimana, Z_i adalah data hasil normalisasi, x_i adalah data asli, \bar{x} adalah nilai rata-rata baku rujukan, dan S adalah nilai simpang baku rujukan. Normalisasi *z-score* dilakukan pada data dari hasil proses transformasi yang disajikan pada Tabel 2.

C. Hasil dan Pembahasan

Hasil analisis wilayah penanaman lada dengan menerapkan metode K-Means telah dilakukan. Setelah mendapatkan data transformasi selanjutnya melakukan proses normalisasi menggunakan Persamaan (7). Hasil normalisasi ditampilkan pada Tabel 3.

Tabel 3. Hasil Normalisasi Metode Z-Score

Kabupaten/ Kota	Luas Lahan (Ha)	Produksi (Ton)	Rata-Rata Produksi (Ton/Ha)	Tenaga Kerja Perkebunan (TKP)
K1	2.431384688	2.726664132	1.173536417	2.116171508
K2	0.477992057	0.039942949	0.339110531	0.698623066
K3	0.407062866	-0.013110227	-0.273686387	0.468775671
K4	-0.812242178	-0.642266455	-1.861433118	-1.082275629
K5	-0.325015385	-0.119405477	0.157522177	0.263481593
K6	-0.682891578	-0.608484146	-0.452410155	-0.749085657
K7	-0.323890293	-0.145302063	0.560561173	-0.261093225

K8	-0.753672746	-0.602291988	-0.047008018	-0.913664498
K9	-0.801767627	-0.618342408	1.494289443	-0.964319524
K10	0.383040196	-0.017404316	-1.090482062	0.423386695

Proses berikutnya, menentukan pusat *cluster* atau *centroid* yang dilakukan secara acak telah dikerjakan. Dalam percobaan ini, penentuan pusat *cluster* untuk $K = 3$ telah dilakukan dengan mengambil data ke-2 sebagai pusat cluster C1, data ke-7 sebagai pusat C2, dan data ke-8 sebagai pusat C3. Hasil penentuan pusat *centroid* awal ditampilkan pada Tabel 4.

Tabel 4. Pusat *Centroid* Awal

Cluster	Jumlah (Ha)	Produksi (Ton)	Rata-Rata Produksi (Ton/Ha)	Tenaga Kerja Perkebunan (TKP)
C1	0.477992057	0.039942949	0.339110531	0.698623066
C2	-0.323890293	-0.145302063	0.560561173	-0.261093225
C3	-0.753672746	-0.602291988	-0.04700802	-0.913664498

Ketiga, penerapan perhitungan jarak *Euclidean Distance* (ED), *Manhattan Distance* (MaN) dan *Minkowski Distance* (MiN) untuk mendapatkan jarak terdekat dimana hasil terbaik menjadi model dalam pengelompokan wilayah penanaman lada. Dalam percobaan ini, perhitungan dengan metode ED menggunakan Persamaan (1) telah menghasilkan 2 iterasi. Metode MaN menggunakan Persamaan (2) telah menghasilkan 4 iterasi. Metode MiN menggunakan Persamaan (3) telah menghasilkan 2 iterasi. Hasil perhitungan pengukuran jarak ditampilkan pada Tabel 5.

Tabel 5. Hasil Perhitungan Jarak 3 Metode

Kabupaten/ Kota	Jarak ke Centroid								
	ED			MaN			MiN		
	C ₁	C ₂	C ₃	C ₁	C ₂	C ₃	C ₁	C ₂	C ₃
K1	3.0245	4.8735	5.8595	0	8.1177	10.972	3.7552	5.5906	7.0636
	92725	66352	97453		34837	22332	57914	88149	54999
K2	0.8705	1.4946	2.4106	6.8920	1.2256	4.0801	1.0594	1.2365	2.6863
	02517	28714	5955	88143	46695	35173	86578	5628	6181
K3	1.0297	1.5867	1.9701	7.8587	0.6836	3.2276	1.2739	1.3041	2.2604
	81581	87965	97585	14822	02963	00344	3127	60766	65729
K4	3.5238	2.7501	1.0895	12.845	4.7282	1.8737	4.4161	2.9790	1.6839
	1127	6281	04258	97412	39287	5089	79496	38799	88349
K5	1.6315	0.8563	1.6460	8.4711	0.7912	2.5010	1.9387	0.3713	1.6759
	9994	11123	31569	73837	73181	49479	21484	511	98312
K6	2.7024	1.3182	0.3794	10.940	2.8228	0.5031	3.3092	1.5471	0.3568
	9897	4213	50259	62828	93444	3442	4536	09883	43865
K7	1.9829	0.2873	1.6271	8.6174	1.7436	2.3547	2.4453	0.3713	1.5039
	57246	1967	9549	81153	58489	42162	76179	196	85957
K8	2.8046	1.0652	0.7401	10.764	2.6754	0.2078	3.3584	1.4452	0.1766
	88136	9354	14834	394	32949	2932	42991	93911	62975
K9	3.2213	1.0924	2.2823	9.9794	4.3315	1.7875	4.0381	1.9994	1.7184
	1657	01753	6977	02913	30737	33356	11483	59934	9638
K10	1.5200	2.1731	1.8779	8.7492	1.4266	3.9706	1.8619	1.9164	2.5526
	09637	62578	36117	16232	9292	90283	97823	572	00454

Keempat, melakukan perhitungan akurasi ketiga jarak dengan menggunakan metode *Silhouette Coefficient* telah dikerjakan. Hasil perhitungan akurasi dengan metode SC dapat dilihat pada Tabel 6.

Tabel 6. Nilai Akurasi Metode SC

Akurasi	Distances
---------	-----------

	ED	MaN*	Min
Silhouette Coefficient	0.1646122	0.286314389	0.028132606

Tabel 6 menunjukkan pengujian akurasi nilai *Silhouette Coefficient* yang paling tinggi adalah metode *Manhattan Distance* dengan nilai 0,286314389. Oleh karena itu, metode ini telah dipilih menjadi model analisa pengelompokan wilayah penanaman lada. Selanjutnya, pengukuran akurasi dengan metode SSE dengan Persamaan (5) dan R^2 dengan Persamaan (6) untuk menguji jumlah *cluster* yang optimal dari metode perhitungan jarak *Manhattan Distance* telah diterapkan. Dimana, pengelompokan dengan 3 *cluster* juga telah dibandingkan dengan jumlah *cluster* sebanyak 4 dan 5 *cluster*. Hasil perhitungan nilai SSE dan R^2 ditampilkan pada Tabel 7.

Tabel 7. Nilai Akurasi Metode SSE dan R^2

Akurasi	Clusters		
	K=3*	K=4	K=5
SSE	238.7377116	301.7159484	343.5910198
R^2	0.459398609	0.408042966	0.416734349

Tabel 7 memperlihatkan bahwa metode SSE dengan nilai *error* terkecil telah diperoleh dengan jumlah 3 *cluster* yang merupakan *cluster* optimal dengan nilai SSE sebesar 238,7377116 dan nilai R^2 sebesar 0,459398609. Dengan kata lain, berdasarkan uji akurasi *cluster* yang optimal telah dilakukan dapat menjadi dasar perhitungan wilayah penanaman lada. Selanjutnya, model *cluster* terbaik tersebut telah diterapkan untuk mengelompokkan wilayah penanaman lada di 10 Kabupaten/Kota. Hasil *centroid* pada iterasi terakhir dan *cluster* ditampilkan pada Tabel 8.

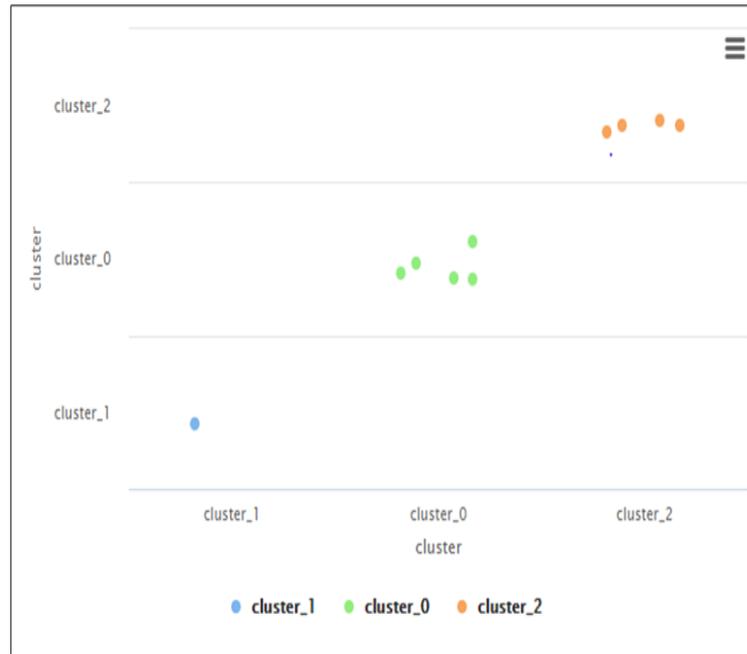
Tabel 8. Hasil *Centroid* dan *Cluster*

Provinsi/ Kabupaten/Kota	Jarak ke Centroid			Hasil Cluster		
	C1	C2	C3	C1	C2	C3
K1	0	8,117734837	10,97222332	*		
K2	6,892088143	1,225646695	4,080135173		*	
K3	7,858714822	0,683602963	3,227600344		*	
K4	12,84597412	4,728239287	1,873750809			*
K5	8,471173837	0,791273181	2,501049479		*	
K6	10,94062828	2,822893444	0,50313442			*
K7	8,617481153	1,743658489	2,354742162		*	
K8	10,764394	2,675432949	0,20782932			*
K9	9,979402913	4,331530737	1,787533356			*
K10	8,749216232	1,426692902	3,970690283		*	

Keterangan:

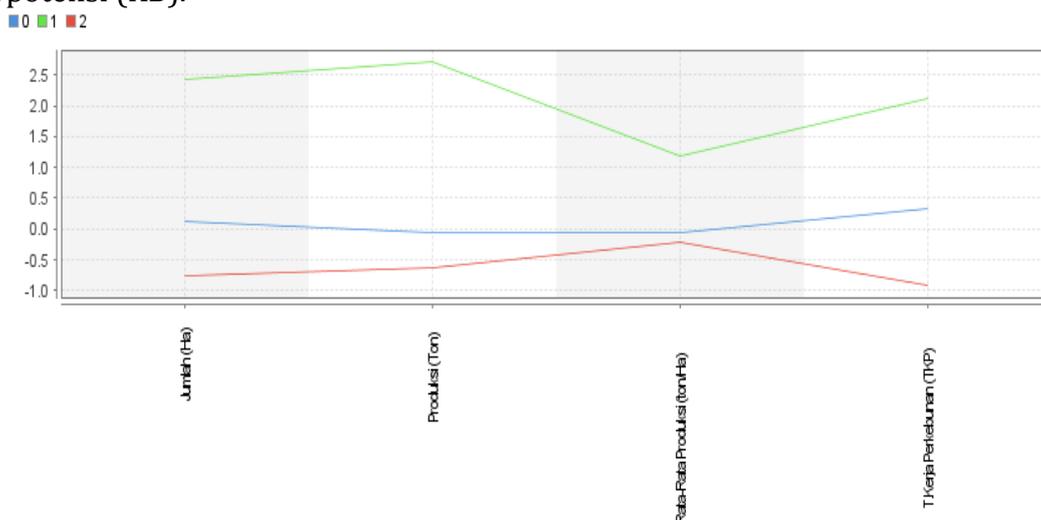
Kabupaten/Kota terdiri dari 10 yaitu (1) K1-Kutai Kartanegara; (2) K2-Kutai Timur; (3) K3-Kutai Barat; (4) K4-Mahakam Ulu; (5) K5-Penajam Paser Utara; (6) K6-Paser; (7) K7-Berau; (8) K8-Samarinda; (9) K9-Balikpapan; (10) K10-Bontang

Hasil proses klusterisasi memperlihatkan wilayah pada *Cluster 1* (C1) sebanyak 1 wilayah yaitu Kutai Kartanegara, *Cluster 2* (C2) sebanyak 5 wilayah yaitu Kutai Timur, Kutai Barat, Penajam Paser Utara (P.U), Berau dan Bontang. Sementara *Cluster 3* (C3) sebanyak 4 wilayah yaitu Mahakam Ulu, Kabupaten Paser, Samarinda dan Balikpapan. Hasil pengelompokan juga ditampilkan dalam bentuk grafik *scatter* menggunakan *software Rapidminer* yang dapat dilihat pada Gambar 1. Hasil pengelompokan menghasilkan *Cluster 0* (C0) berjumlah 5 data, *Cluster 1* (C1) berjumlah 1 data dan *Cluster 2* (C2) berjumlah 4 data.



Gambar 1. Plot Wilayah Penanam Lada Provinsi Kalimantan Timur.

Dalam menentukan *cluster* mana yang termasuk dalam *cluster* dengan kategori wilayah berpotensi baik (BB), wilayah berpotensi sedang (BS) dan wilayah kurang berpotensi (KB) pada *Rapidminer* maka dapat dilihat melalui grafik plot pengelompokan pada Gambar 2. Pada gambar tersebut warna hijau adalah C1, warna biru adalah C0, dan warna merah adalah C2. Dilihat dari grafik Gambar 2, C1 berada pada posisi tertinggi sehingga disimpulkan bahwa C1 merupakan *cluster* dengan kategori wilayah berpotensi baik (BB), C0 berada pada posisi tengah sehingga disimpulkan bahwa C0 merupakan *cluster* dengan kategori wilayah berpotensi sedang (BS), sedangkan C2 berada pada posisi terbawah sehingga disimpulkan bahwa C2 merupakan *cluster* dengan kategori wilayah kurang berpotensi (KB).



Gambar 2. Grafik Plot Pengelompokan.

Berdasarkan hasil pada Gambar 2 dapat diketahui bahwa C1 pada *rapidminer* berkorelasi dengan C1 pada proses perhitungan yang dilakukan dengan wilayah berstatus berpotensi baik (BB) yang terdiri dari 1 wilayah yaitu Kutai Kartanegara, C0 pada *rapidminer* berkorelasi dengan C2 pada proses perhitungan yang dilakukan dengan wilayah berstatus berpotensi sedang (BS) yang terdiri dari 5 wilayah yaitu Kutai Timur, Kutai Barat, Penajam (P.U), Berau dan Bontang dan C2 pada *rapidminer* berkorelasi dengan C3 pada proses perhitungan yang dilakukan dengan wilayah berstatus kurang berpotensi (KB) yang terdiri dari 4 wilayah yaitu Mahakam Ulu, Paser, Samarinda dan Balikpapan.

Hasil pengelompokan tersebut kemudian dibandingkan dengan data pengelompokan dari Dinas Perkebunan Provinsi Kalimantan Timur yang membagi kategori pengelompokan menjadi tertinggi dan terendah berdasarkan hasil produksi. Wilayah kategori penghasil tanaman lada tertinggi terdiri dari 3 daerah yaitu Kutai Kartanegara, Penajam (P.U) dan Berau. Selanjutnya, wilayah kategori terendah terdiri dari 7 wilayah yaitu Kutai Timur, Kutai Barat, Paser, Samarinda, Balikpapan, Mahakam Ulu dan Bontang. Hasil perbandingan tersebut menunjukkan bahwa terdapat perbedaan jumlah kategori dan hasil pengelompokan, dimana penelitian ini menggunakan 3 kategori pengelompokan yang terdiri dari kategori wilayah berpotensi baik, wilayah berpotensi sedang dan wilayah kurang berpotensi, sedangkan Dinas Perkebunan Provinsi Kalimantan Timur menggunakan 2 kategori yang terdiri dari kategori tertinggi dan terendah.

Berdasarkan hasil analisis penelitian yang telah dilakukan, wilayah-wilayah berstatus BB perlu direkomendasikan sebagai wilayah yang perlu mempertahankan dan memaksimalkan penanaman dan produksi dari tanaman lada. Sedangkan, wilayah berstatus BS perlu direkomendasikan untuk lebih meningkatkan kualitas lahan sehingga meningkatkan proses penanaman dan diharapkan nantinya dapat menjadi wilayah yang berpotensi baik (BB). Wilayah yang berstatus KB perlu mendapatkan perhatian dari pemerintah dan pemerintah harus melakukan langkah yang serius untuk memperbaiki wilayah tersebut sehingga mampu mengoptimalkan penanaman dan meningkatkan produksi lada.

D. Simpulan

Penelitian ini telah menerapkan metode *K-Means* untuk pengelompokan wilayah penanam lada khususnya wilayah penanam lada di Provinsi Kalimantan Timur. Berdasarkan uji akurasi menggunakan metode *silhouette coefficient* (SC) diperoleh bahwa untuk studi kasus ini metode pengukuran jarak yang terbaik adalah metode *Manhattan Distance*. Lebih lanjut, berdasarkan uji akurasi *cluster* menggunakan metode *Sum Square Error* (SSE) dan *R-squared* (R^2) diperoleh bahwa 3 *cluster* adalah jumlah *cluster* yang optimal. Hasil tersebut menunjukkan bahwa untuk klusterisasi wilayah penanam tanaman lada lebih cocok menggunakan tiga kategori atau kluster.

E. Ucapan Terima Kasih

Terima kasih ditujukan kepada Dinas Perkebunan Provinsi Kalimantan Timur yang telah membantu penelitian ini.

F. Referensi

- [1] Dinas Perkebunan Provinsi Kalimantan Timur, "Lada," 2020. <https://disbun.kaltimprov.go.id>.
- [2] Dinas Perkebunan Provinsi Kalimantan Timur, "Statistik Perkebunan Provinsi Kalimantan Timur Tahun 2018," 2018. <https://disbun.kaltimprov.go.id>.
- [3] Dinas Perkebunan Provinsi Kalimantan Timur, "Statistik Perkebunan Provinsi Kalimantan Timur Tahun 2016," 2016. <https://disbun.kaltimprov.go.id>.
- [4] S. Sugiartiningsih, "Ekstensifikasi Produksi Lada di Indonesia Kurun Waktu 1971-2015," *Portofolio J. Ekon. Bisnis, Manajemen, Dan Akunt.*, vol. 18, no. 1, pp. 61–72, 2021.
- [5] Y. Sugiyani, "Pengelompokan Wilayah Berdasarkan Potensi Hasil Pertanian Menggunakan Algoritma K-means Di Kota Cilegon," *ProTekInfo (Pengembangan Ris. dan Obs. Tek. Inform.)*, vol. 3, pp. 60–67, 2016.
- [6] H. Andema, S. Defit, and Yuhandri, "Optimalisasi Penggunaan Lahan Perkebunan Kelapa Hibrida Menggunakan K-Means Clustering," *J. Inform. Ekon. Bisnis*, vol. 2, no. 2018, pp. 1–6, 2020, doi: 10.37034/infv2i2.23.
- [7] S. Laia, D. H. Pane, and E. Affandi, "Implementasi Metode K-Means Untuk Mengelompokkan Kawasan Potensi Pertanian Karet Produktif," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 4, p. 282, 2022, doi: 10.53513/jursi.v1i4.5224.
- [8] H. S. Pakpahan, J. A. Widiyans, and H. D. A. Firmada, "Implementasi Metode K-Means Untuk Pengelompokan Potensi Produksi Komoditas Perkebunan," *Adopsi Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 52–60, 2022.
- [9] N. R. Octaviyani, R. Mayasari, and Susilawati, "Implementasi Algoritma K-Means Clustering Status Gizi Balita," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 13, pp. 370–381, 2022.
- [10] U. Ma'rifatin, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien Pada Puskesmas Warujayeng," in *SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 2020, pp. 285–291.
- [11] P. Silitonga and I. S. Morina, "Klusterisasi Pola Penyebaran Penyakit Pasien berdasarkan Usia Pasien dengan Menggunakan K-Means Clustering," *J. TIMES*, vol. 6, no. 2, pp. 22–25, 2017.
- [12] N. Hasanah, M. Ugiarto, and N. Puspitasari, "Sistem pengelompokan curah hujan menggunakan metode k-means di wilayah kalimantan timur," vol. 2, no. 2, 2017.
- [13] C. Astria, A. P. Windarto, A. Wanto, and E. Irawan, "Metode K-Means Pada Pengelompokan Wilayah Pendistribusian Listrik," *Semnar Nas. Sains Teknol. Inf.*, pp. 306–312, 2019.
- [14] D. P. T. Hapsari and E. Widodo, "Pengelompokan Daerah Rawan Kriminalitas di Indonesia Menggunakan Analisis K-Means Clustering," in *Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika dan Nilai-Nilai Islami)*, 2017, vol. 1, no. 1, pp. 147–153.
- [15] N. Dwitri, J. A. Tampubolon, S. Prayoga, F. I. R. H. Zer, and D. Hartama, "Penerapan algoritma K-Means dalam menentukan tingkat penyebaran pandemi COVID-19 di Indonesia," *JurTI (Jurnal Teknol. Informasi)*, vol. 4, no.

- 1, pp. 128–132, 2020.
- [16] M. N. Mara and N. Satyahadewi, “Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis,” *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 02, no. 2, pp. 133–136, 2013.
- [17] Purnawansyah and Havaluddin, “K-Means clustering implementation in network traffic activities,” *Proc. - Cybern. 2016 Int. Conf. Comput. Intell. Cybern.*, pp. 51–54, 2017, doi: 10.1109/CyberneticsCom.2016.7892566.
- [18] K. Singh, D. Malik, and N. Sharma, “Evolving limitations in K-means algorithm in data mining and their removal,” vol. 12, no. April, pp. 105–109, 2011.
- [19] S. Patel and A. Patel, “Performance Analysis and Evaluation of Clustering Algorithms,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 6S2, pp. 179–183, 2019.
- [20] Havaluddin *et al.*, “A Performance Comparison of Euclidean, Manhattan and Minkowski Distances in K-Means Clustering,” *2020 6th Int. Conf. Sci. Inf. Technol. Embrac. Ind. 4.0 Towar. Innov. Disaster Manag. ICSITech 2020*, pp. 184–188, 2020, doi: 10.1109/ICSITech49800.2020.9392053.
- [21] Purnawansyah, Havaluddin, A. F. O. Gafar, and I. Tahyudin, “Comparison Between K-Means and Fuzzy C-Means Clustering in Network Traffic Activities,” *Proc. Elev. Int. Conf. Manag. Sci. Eng. Manag.*, vol. 2, 2018, doi: 10.1007/978-3-319-59280-0.
- [22] Purnawansyah, Havaluddin, A. F. O. Gafar, and I. Tahyudin, “Comparison Between K-Means and Fuzzy C-Means Clustering in Network Traffic Activities,” in *Proceedings of the Eleventh International Conference on Management Science and Engineering Management*, 2017.
- [23] M. Nishom, “Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square,” *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, 2019, doi: 10.30591/jpit.v4i1.1253.
- [24] R. P. Ananto, Y. Purwanto, and A. Noviianty, “Deteksi Jenis Serangan Pada Distributed Denial of Service Berbasis Clustering dan Classification Menggunakan Algoritma Minkowski Weighted K-Means dan Decision Tree,” *e-Proceeding of Engineering*, vol. 4, no. 1, pp. 879–886, 2017.
- [25] B. Wira, A. E. Budianto, and A. S. Wiguna, “Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang,” *RAINSTEK J. Terap. Sains Teknol.*, vol. 1, no. 3, pp. 53–68, 2019, doi: 10.21067/jtst.v1i3.3046.
- [26] B. K. Nelson, “Statistical methodology: V. Time series analysis using autoregressive integrated moving average (ARIMA) models,” vol. 5, no. 7, 1998.
- [27] M. Yoalifa, “Analisis Clustering Data Tingkat Mutu Sekolah Menengah Atas Daerah Kutai Barat Dan Kutai Kartanegara Provinsi Kalimantan Timur Dengan Algoritma K-Means,” 2020.
- [28] S. I. Wardani, W. P. Nurmayanti, and M. Malthuf, “Analisis Cluster Kecamatan di Lombok Timur Berdasarkan Banyaknya Perusahaan dan Cabang Industri,” *J. Appl. Stat. DATA Min.*, vol. 1, no. 2, pp. 1–4, 2020.
- [29] D. Nafkiyah, L. Rifatin, and M. R. Rozikin, “Analisis Cluster dalam Pengelompokan Kabupaten / Kota di Provinsi Jawa Timur berdasarkan

Indikator Pendidikan," *J. Ilm. Mat. dan Pendidik. Mat.*, vol. 12, no. 1, pp. 1-16, 2022.