

Clustering dengan Metode K-Means Berdasarkan Potensi Pertanian di Jawa Tengah

Angeliya Milla Khariyani¹, Evi Fauziyari²

Universitas Negeri Yogyakarta^{1,2}
angeliyamilla.2018@student.uny.ac.id

Abstrak—Sektor pertanian menjadi salah satu kekuatan untuk membangun ekonomi di Indonesia. Saat ini, sektor pertanian menjadi mesin penggerak perekonomian terbesar bagi Indonesia di masa pandemi Covid-19. Salah satu provinsi yang sedang berupaya untuk meningkatkan sektor pertaniannya yaitu Provinsi Jawa Tengah karena di provinsi tersebut memiliki wilayah yang cukup luas. Oleh karena itu untuk meningkatkan sektor pertanian tentunya perlu diketahui daerah-daerah yang memiliki potensi pertanian yang rendah dan daerah-daerah yang memiliki potensi pertanian yang tinggi agar pengelolaan potensi pertanian dapat optimal di setiap daerah. Jika potensi pertanian dapat dikelola dengan baik akan berdampak pada hasil pertanian yang akan meningkat sehingga dapat memulihkan perekonomian Indonesia di masa pandemi Covid-19. Pada penelitian ini terdapat 15 variabel yang menjadi dasar dalam pengelompokan yaitu luas panen, produksi, dan produktivitas pada masing-masing tanaman padi, jagung, kedelai, kacang tanah, dan kacang hijau. Metode *clustering K-Means* digunakan pada penelitian ini. Objek yang mempunyai karakteristik yang sama akan dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang berbeda. Pemilihan jumlah cluster terbaik menggunakan indikator *Davies Bouldin Index*. Dari 35 kabupaten/kota Kota di Jawa Tengah, diperoleh hasil bahwa jumlah *cluster* terbaik yaitu 2 *cluster* berdasarkan nilai *Davies Bouldin Index* yang terkecil dari pengelompokan 2 *cluster*, 3 *cluster* dan 4 *cluster*. Dari hasil tersebut diperoleh bahwa *Cluster 1* merupakan *cluster* dengan kabupaten/kota yang memiliki potensi pertanian yang “rendah” dengan anggota sebanyak 27 kabupaten/kota. Sedangkan, *cluster 2* merupakan *cluster* dengan kabupaten/kota dengan nilai *cluster* yang “tinggi” dengan anggota sebanyak 8 kabupaten/kota.

Kata kunci: *Clustering, K-Means, Potensi Pertanian*

I. PENDAHULUAN

Pangan merupakan hal pokok yang dibutuhkan manusia setiap hari. Pangan sudah menjadi kebutuhan mendasar yang harus dipenuhi setiap saat. Salah satu penyokong utama dalam ketahanan pangan di Indonesia adalah pada sektor pertanian. Sektor pertanian sendiri menjadi salah satu kekuatan untuk membangun ekonomi di Indonesia. Bahkan pada saat membuka Rakornas Pengendalian Inflasi Tahun 2021 Presiden Republik Indonesia, Joko Widodo, mengatakan bahwa sektor pertanian menjadi mesin penggerak perekonomian terbesar bagi Indonesia di masa pandemi Covid-19. Untuk itu perlu dilakukan peningkatan potensi pada sektor pertanian untuk dapat meningkatkan perekonomian Indonesia.

Salah satu provinsi yang sedang mengembangkan ekonominya di sektor pertanian yaitu Provinsi Jawa Tengah. Diketahui bahwa Provinsi Jawa Tengah akan menerapkan konsep pertanian pintar atau *smart farming* yang diharapkan petani-petani di Jawa Tengah dapat meningkatkan efisiensi kerja serta nilai ekonomi dari produk-produk pertanian yang dihasilkan. Pemerintahan Jawa Tengah juga berupaya untuk merevitalisasi sektor usaha dengan perluasan area tanam dan penanaman lahan baru [1]. Untuk peningkatan hal tersebut, tentunya perlu diketahui daerah-daerah yang kurang berpotensi sehingga dapat pemeratakan potensi ekonomi khususnya sektor pertanian di provinsi Jawa Tengah. Hal itu dikarenakan Jawa Tengah memiliki wilayah yang cukup luas. Oleh karena itu, perlu dilakukan pemetaan atau pengelompokan pada potensi pertanian di Jawa Tengah dengan tujuan untuk mengetahui daerah yang memiliki potensi pertanian yang cukup sehingga dapat dimanfaatkan dan juga potensi pertanian yang

kurang sehingga dapat dibenahi. Pemetaan atau pengelompokan tersebut menggunakan metode analisis *cluster*

Analisis *cluster* adalah suatu analisis yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik data. Salah satu metode yang dapat digunakan adalah metode *K-Means*. Tujuannya untuk mengklusterkan data pada *cluster* atau kelompok yang relatif sama. Pemilihan jumlah *cluster* terbaik menggunakan indikator *Davies Bouldin Index*. Berdasarkan uraian di atas, penulis akan melakukan proses pengelompokan kabupaten/kota di Provinsi Jawa Tengah berdasarkan potensi pertanian pada tahun 2019.

II. METODE PENELITIAN

A. Data

Data yang digunakan adalah data-data kabupaten/kota di Jawa Tengah pada tahun 2019. Data yang digunakan merupakan data sekunder yang didapat dari publikasi Buku Jawa Tengah Dalam Angka 2021 yang diperoleh dari *website* Badan Pusat Statistik Jawa Tengah [2].

Variabel yang digunakan dalam penelitian memiliki tipe data numerik. Terdapat 15 variabel yang akan digunakan untuk melakukan proses pengelompokan yang terdapat pada Tabel 1.

TABEL 1. VARIABEL PENELITIAN

Variabel	Keterangan	Deskripsi
X_1	Luas panen padi (m^2)	Luas panen adalah luas tanaman yang dipanen sekaligus/habis/dibongkar dan luas tanaman yang dipanen berkali-kali (lebih dari satu kali)/belum habis.
X_2	Luas panen jagung (m^2)	
X_3	Luas panen kedelai (m^2)	
X_4	Luas panen kacang tanah (m^2)	
X_5	Luas panen kacang hijau (m^2)	
X_6	Produksi padi (ton)	Produksi tanaman adalah hasil perkalian antara luas panen dengan produktivitas (hasil per hektar)
X_7	Produksi jagung (ton)	
X_8	Produksi kedelai (ton)	
X_9	Produksi kacang tanah (ton)	
X_{10}	Produksi kacang hijau (ton)	
X_{11}	Produktivitas padi (kuintal/ha)	Produktivitas adalah kemampuan atau daya dukung lahan pertanian dalam memproduksi tanaman
X_{12}	Produktivitas jagung (kuintal/ha)	
X_{13}	Produktivitas kedelai (kuintal/ha)	
X_{14}	Produktivitas kacang tanah (kuintal/ha)	
X_{15}	Produktivitas kacang hijau (kuintal/ha)	

B. Analisis Cluster

Analisis *cluster* adalah suatu analisis yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik data tersebut. Tujuannya untuk mengklusterkan data pada klaster atau kelompok yang relatif sama. Analisis *cluster* dibagi menjadi 2 metode, yaitu metode *cluster* hirarki dan metode *cluster* nonhirarki. Perbedaan kedua metode tersebut terletak pada penentuan jumlah *cluster* yang dihasilkan. Penentuan jumlah *cluster* pada metode hirarki belum diketahui sebelum memvisualisasikan data dalam bentuk dendrogram, sedangkan metode *cluster* nonhirarki sudah ditentukan sebelumnya [3].

C. Metode K-Means

K-Means clustering adalah salah satu metode *cluster* analisis pada data mining yang termasuk nonhierarki yang bertujuan untuk mempartisi objek yang ada kedalam salah satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama

dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam *cluster* yang berbeda [4]. Dasar algoritma *K-Means* adalah sebagai berikut:

1. Tentukan nilai k sebagai jumlah *cluster* yang ingin dibentuk
2. Bangkitkan k *centroid* (titik pusat *cluster*) awal secara/acak, kemudian untuk menghitung *centroid cluster* k -I berikutnya, digunakan rumus sebagai berikut:

$$v = \sqrt{\frac{\sum_{i=1}^n x_i}{n}} \quad (1)$$

dengan :

i : 1,2,...,n

v : *centroid* pada *cluster*

x_i : objek ke- i

n : banyaknya objek/ jumlah objek yang menjadi anggota *cluster*

3. Hitung jarak setiap data ke masing-masing *centroid* menjadi anggota *cluster*

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

dimana:

d_{ij} : jarak objek antara objek x dan y

n : jumlah atribut

x_i : objek data

y_i : data *cluster*

4. Alokasikan masing-masing objek kedalam *centroid* yang paling terdekat
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan (1)
6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama.

Adapun karakteristik dari algoritma *K-Means* salah satunya adalah sangat sensitif dalam penentuan titik pusat awal *cluster* karena *K-Means* membangkitkan titik pusat *cluster* awal secara *random*. Pada saat pembangkitkan awal titik pusat yang *random* tersebut mendekati solusi akhir pusat *cluster*, *K-Means* mempunyai posibilitas yang tinggi untuk menemukan titik pusat *cluster* yang tepat. Sebaliknya, jika awal titik pusat tersebut jauh dari solusi akhir pusat *cluster*, maka besar kemungkinan ini menyebabkan hasil pengelompokan yang tidak tepat. Akibatnya *K-Means* tidak menjamin hasil pengklasteran yang unik. Inilah yang menyebabkan metode *K-Means* sulit untuk mencapai optimum global, akan tetapi hanya minimum lokal. Selain itu algoritma *K-Means* hanya bisa digunakan untuk data yang atributnya bernilai numerik.

D. Davies Bouldin Index

Davies Bouldin Index merupakan salah satu metode evaluasi internal yang mengukur evaluasi *cluster* pada suatu metode pengelompokan yang didasarkan pada nilai kohesi dan separasi. Dalam suatu pengelompokan, kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap *centroid* dari yang diikuti. Sedangkan separasi didasarkan pada jarak antar *centroid* dari *clusternya*. *Sum of square within cluster* (SSW) merupakan persamaan yang digunakan untuk mengetahui kohesi dalam sebuah *cluster* ke- i yang dirumuskan sebagai berikut:

$$SSW = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \tag{3}$$

Dari persamaan tersebut, m_i merupakan jumlah data dalam *cluster* ke- i , c_i adalah *centroid cluster* ke- i , dan $d()$ merupakan jarak setiap data ke *centroid* yang dihitung menggunakan jarak Euclidean.

Sum of square between cluster (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan:

$$SSB_{i,j} = d(c_i, c_j) \tag{4}$$

Setelah nilai kohesi dan separasi diperoleh, kemudian dilakukan pengukuran rasio (R_{ij}) untuk mengetahui nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j . *Cluster* yang baik adalah *cluster* yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Nilai rasio dihitung menggunakan persamaan sebagai berikut:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \tag{5}$$

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *Davies Bouldin Index* dari persamaan berikut ini:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \tag{6}$$

Dari persamaan tersebut, k merupakan jumlah *cluster* yang digunakan. Semakin kecil nilai *Davies Bouldin Index* yang diperoleh ($\text{non-negatif} \geq 0$), maka semakin baik *cluster* yang diperoleh dari pengelompokan *K-Means* yang digunakan [5].

III. HASIL DAN PEMBAHASAN

A. Karakteristik Data

Statistik deskriptif digunakan untuk mengetahui karakteristik dari data potensi pertanian di Provinsi Jawa Tengah. Statistik deskriptif untuk variabel penelitian disajikan dalam bentuk tabel, seperti pada Tabel 2.

TABEL 2. KARAKTERISTIK DATA

Variabel	Min	Max	Mean	Sum
X_1	45,27	136209,6	47956,55	1678479
X_2	0	116498	16193,34	566767
X_3	0	7495	1084,143	37945
X_4	0	24802	1545,6	54096
X_5	0	28290	2413,886	84486
X_6	288,24	772521,5	275875,8	9655654
X_7	0	737183	99066,14	3467315
X_8	0	13961	1837,8	64323
X_9	0	29651	2131,543	74604
X_{10}	0	37636	2856,914	99992
X_{11}	44,04	69,19	56,87371	1990,58

X_{12}	0	90,7	55,42743	1939,96
X_{13}	0	23,26	13,64229	477,48
X_{14}	0	25,42	14,53657	508,78
X_{15}	0	15,06	6,242571	218,49

Pada tabel 2, dapat dilihat bahwa terdapat 5 jenis tanaman yang menjadi variabel dalam penelitian ini. Berdasarkan rata-rata dari semua kabupaten/kota yang ada di Jawa Tengah, padi menjadi jenis tanaman dengan luas panen, produksi dan produktivitas yang paling tinggi nilainya dibandingkan dengan keempat jenis tanaman yang lain. Wilayah dengan luas panen padi tertinggi yaitu Kabupaten Grobogan sebesar 136209,59 m². Sedangkan wilayah dengan luas panen terendah yaitu Kota Surakarta sebesar 45,27 m². Rata-rata luas tanaman terendah yaitu tanaman kedelai sebesar 1084,143 hektar. Sedangkan untuk produktivitas tanaman terendah yaitu tanaman kacang hijau sebesar 6,24 kwintal/ha.

B. Penentuan Jumlah Cluster

Cluster model yang dihasilkan dari pengelompokan data berdasarkan potensi pertanian dengan algoritma *K-Means* ditampilkan dalam tabel. Pada tabel tersebut terlihat bahwa jumlah cluster yang digunakan sebanyak 2 dengan keanggotaan cluster 1 terdiri dari 27 wilayah, cluster 2 terdiri dari 8 wilayah dengan jumlah keseluruhannya adalah 35 wilayah kabupaten/kota di Jawa Tengah. Pada clustering dengan jumlah cluster sebanyak 3 keanggotaan yang dihasilkan yaitu cluster 1 terdiri dari 16 wilayah, cluster 2 terdiri dari 6 wilayah dan cluster 3 terdiri dari 13 wilayah. Sedangkan pada jumlah cluster yang digunakan sebanyak 4 dengan keanggotaan cluster 1 terdiri dari 6 wilayah, cluster 2 terdiri dari 13 wilayah, cluster 3 terdiri dari 9 wilayah dan cluster 4 terdiri dari 7 wilayah.

TABEL 3. DAVIES BOULDIN INDEX MASING-MASING CLUSTER

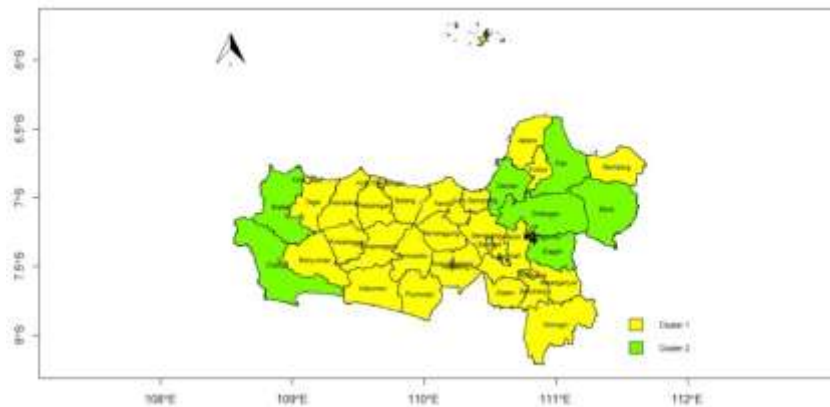
Jumlah Cluster	Hasil Cluster	Performance Vektor (Davies Bouldin Index)
2 cluster	Cluster 1 = 27 wilayah	0,837
	Cluster 2 = 8 wilayah	
3 cluster	Cluster 1 = 16 wilayah	0,851
	Cluster 2 = 6 wilayah	
	Cluster 3 = 13 wilayah	
4 cluster	Cluster 1 = 6 wilayah	0,963
	Cluster 2 = 13 wilayah	
	Cluster 3 = 9 wilayah	
	Cluster 4 = 7 wilayah	

Dari hasil analisis yang dilakukan cluster terbaik yang dapat dibentuk yaitu 2 cluster karena Davies Bouldin Index yang dihasilkan paling kecil dibanding dengan 2 cluster yang lain. Davies Bouldin Index yang dihasilkan pada clustering dengan 2 cluster yaitu 0,837 (paling mendekati noll).

C. Clustering dengan Metode K-Means

Hasil clustering diperoleh dari perhitungan menggunakan software R dengan metode *K-Means*. Berikut hasil clustering dengan metode *K-Means* dengan cluster yang digunakan sebanyak 2:

1. Cluster 1 terdiri dari 27 kabupaten/kota, yaitu: Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Sukoharjo, Karanganyar, Rembang, Kudus, Jepara, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, dan Kota Tegal
2. Cluster 2 terdiri dari 8 kabupaten/kota, yaitu: Cilacap, Wonogiri, Sragen, Grobogan, Blora, Pati, Demak dan Brebes



GAMBAR 1. PETA CLUSTERING KABUPATEN/KOTA DI JAWA TENGAH BERDASARKAN POTENSI PERTANIAN

D. Karakteristik Cluster

Dari perhitungan sebelumnya, diperoleh 2 cluster. Cluster 1 terdiri dari 27 kabupaten/kota, sedangkan cluster 2 terdiri dari 8 kabupaten/kota. Untuk mengetahui karakteristik masing-masing cluster digunakan tabel statistik deskriptif masing-masing cluster.

TABEL 4. KARAKTERISTIK CLUSTER 1

Variabel	Min	Max	Mean	Sum
X_1	45,27	76154,59	32391,8	874578,8
X_2	0	34968	8664,7	233946
X_3	0	2135	511,9	13822
X_4	0	3186	731,9	19760
X_5	0	10086	755,9	20409
X_6	288,24	427165,3	180356,2	4869618
X_7	0	215858	51373,0	1387071
X_8	0	3410	840,6	22697
X_9	0	4753	1047,9	28292
X_{10}	0	6517	677,1	18283
X_{11}	44,04	69,19	56,2	1518,15
X_{12}	0	90,7	53,3	1437,78
X_{13}	0	21,82	12,4	335,43
X_{14}	0	20,57	13,6	365,93
X_{15}	0	14,57	4,4	117,76

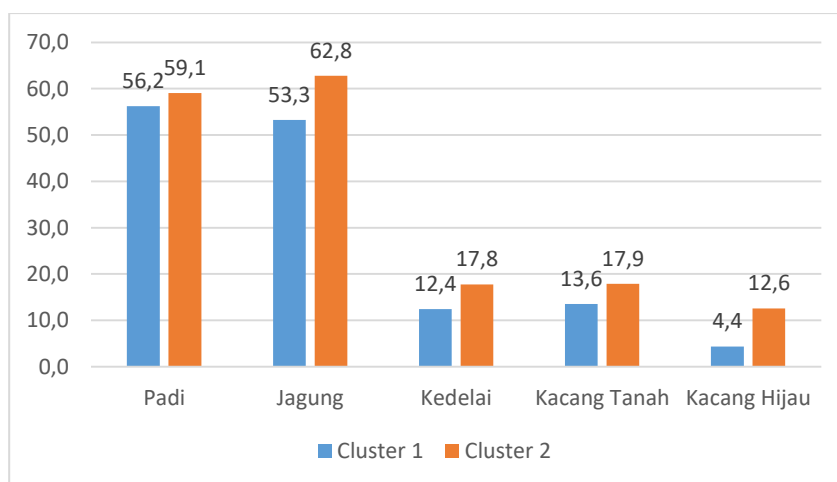
Dari tabel 4, dapat dilihat bahwa semua variabel kecuali luas panen padi (X_1), produksi padi (X_6), dan produktivitas padi (X_{11}) memiliki nilai 0. Artinya pada cluster 1 terdapat kabupaten/kota yang tidak memiliki hasil panen tanaman pertanian. Nilai maksimum dan rata-rata tertinggi terdapat pada variabel X_6 yaitu produksi padi dengan nilai 427165,3 dan 180356,2.

TABEL 5. KARAKTERISTIK CLUSTER 2

Variabel	Min	Max	Mean	Sum
X_1	63109,27	136209,6	100487,6	803900,5
X_2	7876	116498	41602,63	332821

X_3	1264	7495	3015,375	24123
X_4	75	24802	4292	34336
X_5	38	28290	8009,625	64077
X_6	341370,3	772521,5	598254,5	4786036
X_7	44498	737183	260030,5	2080244
X_8	2772	13961	5203,25	41626
X_9	126	29651	5789	46312
X_{10}	45	37636	10213,63	81709
X_{11}	53,09	68,66	59,05375	472,43
X_{12}	55,87	69,5	62,7725	502,18
X_{13}	13,21	23,26	17,75625	142,05
X_{14}	11,95	25,42	17,85625	142,85
X_{15}	10,12	15,06	12,59125	100,73

Dari tabel 5, terlihat bahwa nilai minimum dari semua variabel memiliki nilai. Nilai maksimum dan rata-rata tertinggi juga terdapat pada variabel X_6 atau produksi padi yaitu sebesar 772521,5 dan 598254,5. Nilai tersebut lebih tinggi dari *cluster* 1. Secara keseluruhan dapat dilihat bahwa *cluster* 2 memiliki nilai yang lebih tinggi dari *cluster* 1. Untuk melihat lebih jelas digunakan gambar grafik untuk melihat perbandingan kedua *cluster*, seperti pada gambar 2.



GAMBAR 2. GRAFIK RATA-RATA PRODUKTIVITAS KELOMPOK CLUSTER

Pada gambar 1 tampak *cluster* 2 memiliki nilai rata-rata produktivitas yang lebih tinggi dibandingkan *cluster* 1. Tanaman jagung menjadi nilai rata-rata produktivitas tertinggi pada *cluster* 2. Sedangkan pada *cluster* 1, tanaman padi merupakan nilai yang tertinggi.

IV. SIMPULAN DAN SARAN

Dari hasil penelitian yang dilakukan dapat disimpulkan bahwa penerapan metode *K-Means* untuk *clustering* wilayah kabupaten/kota di Jawa Tengah berdasarkan potensi pertanian berhasil diterapkan sehingga dapat dilakukan pengelolaan potensi pertanian dengan optimal sesuai sasaran. Pengelompokan ini terdiri dari 35 kabupaten/kota kemudian didapatkan jumlah *cluster* terbaik menggunakan *Davies Bouldin Index* sebanyak 2 *cluster* dengan nilai 0,837. Hasil pengelompokan terdapat 2 *cluster* yang terdiri dari *cluster* 1 dengan jumlah keanggotaan 27 kabupaten/kota dan *cluster* 2 dengan jumlah keanggotaan 8 kabupaten/kota. *Cluster* 1 menyatakan potensi pertanian yang “rendah” dan *cluster* 2 menyatakan potensi pertanian yang “tinggi”.

UCAPAN TERIMA KASIH

Ucapan terima kasih, kami sampaikan kepada pihak-pihak yang telah membantu dalam proses pembuatan penelitian ini. Pihak Universitas Negeri Yogyakarta yang telah menyediakan sarana dan prasarana keilmuan statistika dan juga Badan Pusat Statistik yang telah menyediakan data penelitian ini. Semoga senantiasa diberikan kemudahan dan kelancaran dalam setiap langkah yang diambil.

DAFTAR PUSTAKA

- [1] Ikhsan, M Faisal Nur. “Sektor Pertanian Jateng Masih Menjanjikan”. *Bisnis.cdm*, <https://semarang.bisnis.com/read/20211228/535/1482626/sektor-pertanian-jateng-masih-menjanjikan>. Di akses pada 2 Feb. 2022.
- [2] BPS. *Provinsi Jawa Tengah Dalam Angka 2021*. Badan Pusat Statistik Jawa Tengah, <https://www.bps.go.id/>. Di akses pada 30 Jan. 2022.
- [3] Jannah, U. “Perbandingan Jarak Euclid dengan Jarak Mahalanobis pada Analisis Cluster Hirarki”. Universitas Islam Negeri Maulana Malik Ibrahim : Malang, 2010.
- [4] Ediyanto, Mara, M. N., & Setyahadewi, N. “Pengklasifikasian Karakteristik dengan Metode K-means Cluster Analysis”. *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*, 2013, 133-136.
- [5] Prasetyo, E. *Data Mining Mengolah Data menjadi Informasi dengan Matlab*. Yogyakarta: Andi, 2016.