

# Klastering Sayuran Unggulan Menggunakan Algoritma *K-Means*

<http://dx.doi.org/10.28932/jutisi.v8i3.5277>

Riwayat Artikel

Received: 15 Agustus 2022 | Final Revision: 05 Desember 2022 | Accepted: 05 Desember 2022

Creative Commons License 4.0 (CC BY – NC)



Lina Mardiana Harahap<sup>✉#1</sup>, Wahyu Fuadi<sup>#2</sup>, Lidya Rosnita<sup>#3</sup>, Eva Darnila<sup>#4</sup>, Rini Meiyanti<sup>#5</sup>

<sup>#</sup> Program Studi Teknik Informatika, Universitas Malikussaleh

Jalan Batam, Blang Pulo, Muara Satu, Lhokseumawe, Aceh, 24352, Indonesia

<sup>1</sup>lina.180170126@mhs.unimal.ac.id

<sup>2</sup>wahyu.fuadi@unimal.ac.id

<sup>3</sup>lidyarosnita@unimal.ac.id

<sup>4</sup>eva.darnila@unimal.ac.id

<sup>5</sup>rinimeiyanti@unimal.ac.id

<sup>✉</sup>Corresponding author: <sup>1</sup>lina.180170126@mhs.unimal.ac.id

**Abstrak** — Hortikultura khususnya sayuran memiliki potensi besar untuk dikembangkan karena menjadi sumber pendapatan bagi masyarakat dan petani kecil di setiap daerah disebabkan negara Indonesia disebut sebagai negara agraris dengan sebagian besar bekerja dibidang pertanian. Kabupaten Mandailing Natal merupakan kabupaten dengan wilayah terluas di provinsi Sumatera Utara tapi Mandailing Natal belum dapat mengungguli produksi panen tanaman sayuran di Sumatera Utara. Metode penambangan data dapat menemukan pola yang menarik dan tidak terlihat dalam kumpulan data salah satu metodenya adalah algoritma K-Means klastering yang mengelompokkan data ke dalam cluster berdasarkan kesamaan karakteristik data. Pada penelitian ini dilakukan klastering pada data sayuran yang bertujuan untuk mengetahui komoditi yang berpotensi pada setiap daerah di Kabupaten Mandailing Natal, tanaman yang berpotensi di daerah akan tetap dijaga dan ditingkatkan produksinya sedangkan tanaman sayuran yang produksinya masih rendah akan menjadi prioritas untuk meningkatkan hasil produksinya. Metode penelitian yang digunakan pada penelitian ini dengan melakukan pengumpulan data sayuran dari Badan Pusat Statistik berupa data luas panen, produksi, luas tanaman, dan luas penanaman baru. Selain itu, pengumpulan data juga dilakukan dengan melakukan kajian teori pada jurnal. Hasil klastering sayuran unggulan menggunakan algoritma K-Means berupa pengelompokan potensi ke dalam 3 klaster yaitu klaster rendah, sedang, dan tinggi dan didapatkan output berupa sistem berbasis web dalam pengaplikasiannya. Adapun hasil analisa klastering yang didapatkan dengan masing-masing total data 69 data yaitu cabai besar dengan hasil C1 81%, C2 16% dan C3 3%. Cabai Rawit C1 29%, C2 48% dan C3 23%. Kacang Panjang C1 26%, C2 38% dan C3 36%. Kangkung C1 39%, C2 36% dan C3 25%. Terung C1 43%, C2 29% dan C3 28%. Tomat C1 41%, C2 58% dan C3 1%.

**Kata kunci**—Data; Hortikultura; Klastering; K-Means.

## *Clustering of Featured Vegetables Using the K-Means Algorithm*

**Abstract** — *Horticulture, especially vegetables, has great potential to be developed because it becomes a source of income for the community and small farmers in each region because Indonesia is called an agrarian country with most of them working in*

agriculture. Mandailing Natal Regency is the district with the largest area in North Sumatra province, but Mandailing Natal has not been able to outperform vegetable crop production in North Sumatra. Data mining methods can find interesting and invisible patterns in data sets. One of the methods is the K-Means clustering algorithm which groups data into clusters based on the similarity of data characteristics. In this study, vegetable data was clustered which aims to determine the potential commodities in each area in Mandailing Natal Regency, plants that have potential in the area will be maintained and their production increased, while vegetable crops whose production is still low will be a priority to increase their production. The research method used in this study was to collect vegetable data from the Badan Pusat Statistik in the form of data on harvested area, production, plant area, and new planting area. In addition, data collection was carried out by conducting theoretical studies in journals. The results of clustering superior vegetables using the K-Means Algorithm are in the form of potential grouping into 3 clusters, namely low, medium, and high clusters and the output is a web-based system in its application. The results of the clustering analysis were obtained with each total data of 69 data, namely big chili with C1 81%, C2 16% and C3 3%. Cayenne C1 29%, C2 48% and C3 23%. Long Beans C1 26%, C2 38% and C3 36%. Kale C1 39%, C2 36% and C3 25%. Eggplant C1 43%, C2 29% and C3 28%. Tomato C1 41%, C2 58% and C3 1%.

Keywords— Clustering; Data; Horticulture; K-Means

## I. PENDAHULUAN

Hortikultura memiliki potensi besar untuk dikembangkan sehingga dapat mendukung ketahanan pangan nasional. Kontribusi hortikultura terhadap perekonomian Indonesia didasarkan pada Produk Domestik Bruto (PDB) yang merupakan salah satu indikator ekonomi makro yang menentukan peran dan kontribusi subsektor hortikultura terhadap pendapatan nasional. Pengembangan tanaman hortikultura merupakan salah satu agribisnis yang bernilai ekonomis dan berpotensi menjadi sumber pendapatan bagi masyarakat dan petani kecil di setiap daerah [1].

Tanaman hortikultura terutama sayur-sayuran merupakan bahan pangan asal tumbuhan yang mengandung kadar air tinggi dan dapat dikonsumsi dalam keadaan segar atau diolah. Selain itu sayuran merupakan salah satu sumber pangan yang memberikan banyak manfaat untuk dikonsumsi oleh masyarakat Indonesia. Oleh sebab itu, mengkonsumsi sayuran menjadi salah satu upaya untuk menjaga kesehatan [2].

Indonesia memiliki tanah yang subur sehingga banyak berbagai jenis tanaman yang dapat tumbuh dengan cepat. Hal itu disebabkan posisi geografis Indonesia yang sangat strategis yaitu terletak pada daerah tropis yang memiliki curah hujan tinggi. Maka dari itu, Indonesia disebut sebagai negara agraris yaitu negara dengan sebagian besar penduduknya bekerja di bidang pertanian.

Provinsi Sumatera Utara tergolong ke dalam daerah beriklim tropis, karena terletak dekat garis khatulistiwa yaitu  $1^{\circ}$ - $4^{\circ}$  ° Lintang Utara dan  $98^{\circ}$ - $100^{\circ}$  Bujur Timur. Sumatera Utara memiliki luas daratan sekitar 72.981,23 Km persegi dengan jumlah kabupaten dan kota adalah sebanyak 33. Kabupaten Mandailing Natal merupakan kabupaten dengan wilayah terluas di provinsi Sumatera Utara yaitu sebanyak 6.620,70 km<sup>2</sup> atau 8,04 % dari total wilayah Sumatera Utara. Lapangan usaha pada bidang pertanian di Mandailing Natal menyumbang Produk Domestik Regional Bruto terbesar pada tahun 2020. Lapangan usaha pertanian memberi kontribusi sebesar 43,55 persen terhadap PDRB ADHB [3].

Metode penambangan data dapat menemukan pola yang menarik dan tidak terlihat dalam kumpulan data. Metode data mining juga dapat digunakan untuk mengatasi kesalahan pada user dalam pengolahan data [4]. Salah satu metode data mining yang dapat digunakan untuk memetakan atau mengelompokkan data sejenis adalah klastering. Klastering memiliki keunggulan dibandingkan metode data mining lainnya karena dapat mengklasifikasikan data tanpa pengetahuan sebelumnya. *Clustering* membagi data dan mengelompokkan data ke dalam cluster berdasarkan kesamaan tipe data. Teknik *clustering* mencakup beberapa algoritma, salah satu yang paling populer adalah *K-Means* [5].

Algoritma *K-Means* telah diterapkan dalam penyelesaian masalah di berbagai bidang. Dibidang transportasi, algoritma *K-Means* digunakan untuk mengelompokkan kendaraan bermotor menurut kegunaannya [6]. Dibidang kependudukan, algoritma *K-Means* digunakan untuk mengelompokkan data kemiskinan pada provinsi Banten [7]. Pada bidang pariwisata, pengelompokan data jumlah pengunjung setiap tahun berdasarkan objek wisata di Jakarta [8].

Pada penelitian ini dilakukan klastering pada data sayuran unggulan di setiap kecamatan yang ada di Kabupaten Mandailing Natal berdasarkan data hasil potensi pertanian yang dimiliki oleh setiap kecamatan dengan menggunakan metode Algoritma *K-Means* agar dapat digunakan oleh pemerintah daerah dan juga masyarakat setempat. Adapun yang melatar belakangi penelitian ini menggunakan algoritma *K-Means* yaitu pada penelitian terdahulu yang telah berhasil dilakukan dapat disimpulkan bahwa hasil evaluasi *cluster* untuk algoritma *K-Means* memiliki nilai validitas DBI yang jauh lebih rendah dibandingkan dengan nilai DBI algoritma *K-Medoids* [9]. Selain itu pada penelitian terdahulu diketahui bahwa pengelompokan menggunakan metode *K-Means* lebih baik dibandingkan dengan metode *SOM* [10].

Tujuan dari penelitian ini adalah untuk mengetahui komoditi yang berpotensi pada setiap daerah di Kabupaten Mandailing Natal, tanaman yang berpotensi di daerah akan tetap dijaga dan ditingkatkan produksinya sedangkan tanaman sayuran yang produksinya masih rendah akan menjadi prioritas untuk meningkatkan hasil produksinya.

Penetapan komoditas unggulan di suatu wilayah menjadi suatu keharusan. Pemerintah Daerah perlu menentukan sektor dan komoditi apa saja yang diperkirakan bisa tumbuh cepat di wilayah tersebut. Sektor dan komoditi tersebut haruslah yang

merupakan sektor unggulan atau mempunyai prospek untuk dipasarkan ke luar wilayah atau diekspor di masa yang akan datang dan dapat dikembangkan secara maksimal [2].

Sektor pertanian terdiri dari tujuh subsektor pertanian sehingga perlu untuk mengetahui potensi komoditas di suatu wilayah agar diketahui komoditas yang paling unggul dan dapat dikembangkan dengan cepat, baik karena potensi alam maupun karena sector tersebut memiliki keunggulan komparatif untuk dikembangkan. Jadi, dengan modal yang sama, dapat berproduksi dalam waktu yang relatif singkat dan volume sumbangan untuk perekonomian daerah juga [11].

## II. METODE PENELITIAN

### A. Klastering

Klastering adalah proses membuat pengelompokan sehingga semua anggota dari setiap partisi mempunyai persamaan berdasarkan matriks tertentu. Analisis klaster atau analisis kelompok merupakan teknik analisis data yang bertujuan untuk mengelompokkan individu atau objek ke dalam beberapa kelompok yang memiliki sifat berbeda antar kelompok, sehingga individu atau objek yang terletak di dalam satu kelompok akan mempunyai sifat relatif homogen [12].

Klastering juga dapat diartikan metode segmentasi data yang diimplementasikan dalam beberapa bidang, diantaranya marketing, analisa masalah bisnis segmentasi pasar dan prediksi, pola dalam bidang computer vision, zonasi wilayah hingga identifikasi objek dan pengolahan citra [13]. Klastering digunakan untuk mencapai berbagai tujuan seperti pada penelitian [14] di masa pandemi klastering dapat membantu memetakan daerah yang berpotensi dalam penyebaran virus Corona yang bertujuan untuk mengetahui daerah yang beresiko pandemi Covid-19 di Indonesia.

### B. Algoritma K-Means

*K-Means* termasuk algoritma *clustering* dengan proses berulang-ulang. Huruf *K* diartikan sebagai jumlah *cluster* yang hendak dibuat. Selanjutnya nilai *K* ditetapkan secara acak. Sedangkan *means* adalah nilai sementara yang menjadi pusat dari klaster atau disebut juga dengan *centroid* [7].

*K-Means* adalah teknik klastering data *non-hierarchical* yang berusaha membagi data yang ada menjadi satu atau lebih *cluster* atau grup, dimana data dengan karakteristik yang sama dikelompokkan ke dalam klaster yang sama dan data dengan karakteristik yang berbeda dikelompokkan ke dalam grup yang berbeda [15].

*K-Means* sering digunakan karena memiliki kemampuan untuk mengelompokkan data dalam jumlah besar dengan waktu komputasi yang relatif cepat dan efisien. Namun kelemahan dari metode ini adalah kelemahan dalam menganalisis sebaran data dan bergantung pada inisialisasi *centroid*. *K-Means* hanya melihat *range* data pada setiap *centroid* pada setiap cluster [16].

Menurut Govender & Sivakumar (2020) dalam jurnal [17] kelebihan algoritma *K-Means* adalah sebagai berikut:

1. Kompleksitas Rendah
2. Perhitungannya cepat
3. Dapat menangani data yang besar
4. Anggota dalam klaster dapat disesuaikan

### C. Euclidean Distance

*Euclidean distance* merupakan salah satu metode perhitungan jarak yang digunakan untuk mengukur jarak dari 2 (dua) buah titik dalam *euclidean space* meliputi bidang *euclidean* dua dimensi, tiga dimensi, atau bahkan lebih [13]. Adapun rumus untuk menghitung jarak *euclidean* adalah sebagai berikut:

$$D(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

Dimana  $d$  = titik dokumen

$x_i$  = data kriteria

$\mu_j$  = *centroid* pada *cluster* ke- $j$ s

### D. Centroid

*Centroid* merupakan titik pusat yang ditetapkan secara acak atau random. Pada perhitungan titik pusat baru pada iterasi kedua dan seterusnya menggunakan rumus sebagai berikut:

$$y_i = \frac{\sum_{i=0}^n x_i}{n} \quad (2)$$

Keterangan:

$y_i$  = centroid pada kluster

$x_i$  = objek pengamatan ke  $i$

$n$  = banyaknya objek yang menjadi anggota kluster

#### E. Tahapan Algoritma K-Means

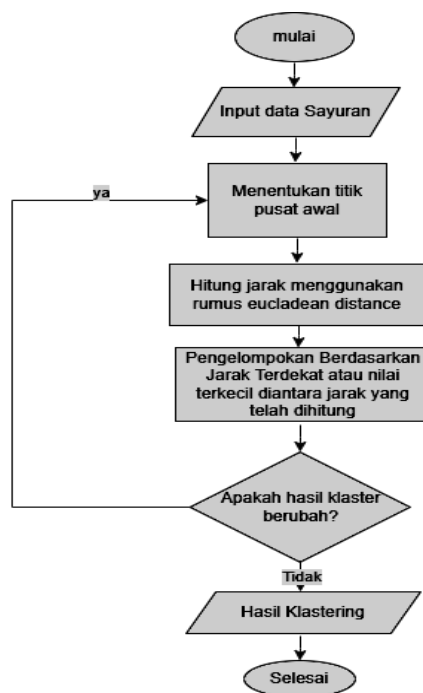
Metode *K-Means Cluster Analysis* menggunakan cara kerja algoritma sebagai berikut:

1. Tentukan  $K$  sebagai jumlah kluster yang terbentuk dengan mempertimbangkan teori atau konsep yang relevan sehingga dapat disepakati berapa banyak kluster yang ingin dibentuk.
2. Ambil sebanyak  $K$  titik pusat kluster. Penentuan titik pusat ini dapat dilakukan secara acak dari data untuk pertama kali, dan rumus berikut digunakan untuk menghitung titik pusat kluster berikutnya menggunakan rumus pada persamaan (2):
3. Menghitung jarak menggunakan *euclidean distance* dengan rumus pada persamaan (1).
4. Setiap objek dinyatakan sebagai anggota kluster tersebut jika jaraknya terdekat dengan titik pusat kluster. Kemudian tentukan posisi centroid baru dengan menggunakan persamaan (2).
5. Kembali ke langkah 3 jika posisi *centroid* baru tidak sama.

Pemeriksaan konvergensi dilakukan dengan membandingkan matriks kelompok dalam iterasi sebelumnya dengan elemen matriks kelompok selama pengulangan. Jika hasilnya sama, algoritma *K-Means* selesai, tetapi jika berbeda, artinya belum konvergen, sehingga perlu dilakukan pengulangan sampai konvergen oleh Johnson et al., (2007) dalam [18].

#### F. Skema K-Means

Skema sistem dapat dilihat pada gambar 1.



Gambar 1. Skema k-means

Gambar 1 merupakan gambaran proses atau skema perancangan sistem pada penelitian ini, dari tahap mulai sampai selesai:

1. Proses skema algoritma *K-Means* dimulai.
2. Proses penginputan data sayuran unggulan yaitu cabai besar, cabai rawit, kacang panjang, tomat, terung dan kangkung.
3. Menentukan titik pusat awal yang diinginkan dengan cara mengurutkan data dari yang terkecil hingga terbesar pada setiap variabel yang dimana banyak seluruh data yaitu sebanyak 69 data dibagi dengan jumlah kelompok sebanyak 3 sehingga pengambilan rata-ratanya memiliki rentang 23 pada masing-masing kluster. Kemudian untuk memilih masing-masing titik pusat kelompok dengan cara:  
C1 (rendah) = nilai rata-rata dari jumlah data ke-1 sampai data-ke 23.

- C2 (sedang) = nilai rata-rata dari jumlah data ke-24 sampai data-ke 46  
C3 (tinggi) = nilai rata-rata dari jumlah data ke-47 sampai data-ke 69.
4. Setelah menentukan titik pusat awal maka dilakukan perhitungan jarak *euclidean* dengan rumus pada persamaan (1).
  5. Pengelompokan dilakukan berdasarkan jarak terdekat atau jarak terkecil dari hasil perhitungan jarak *euclidean* pada tiap data.
  6. Hasil pengelompokan menggunakan *K-Means* berhenti apabila hasil klastering tidak berubah. Apabila hasil klastering berubah maka menghitung titik pusat baru dengan rumus pada persamaan (2). Kemudian menghitung jarak *euclidean* kembali dan mengelompokkan berdasarkan jarak terkecil.
  7. Jika perhitungan sudah selesai maka akan keluar hasil klaster yang dapat menentukan data tersebut masuk pada klaster rendah, sedang atau tinggi.
  8. Proses perhitungan selesai.

#### G. Silhouette Coefficient

*Silhouette Coefficient* adalah metode evaluasi *cluster* yang menggabungkan metode *cohesion* dengan *separation*. *Cohesion* diukur dengan menghitung semua objek yang ada dalam sebuah *cluster* dan *separation* diukur dengan menghitung jarak rata-rata setiap objek dalam sebuah *cluster* dengan *cluster* paling dekat [19]. Nilai *silhouette* untuk keseluruhan data dengan jumlah *cluster*  $k$ , dapat didefinisikan sebagai  $sil(k)$  yang dihitung dengan persamaan (3) yakni rata-rata *silhouette value* untuk semua *cluster* [20].

$$sil(c) = \frac{sil(k)}{k} / \sum sil(ci) \quad (3)$$

Keterangan:

$sil(k)$  : nilai *silhouette* semua *cluster*

$|k|$  : banyaknya *cluster*  $k$

$sil(ci)$  : rata-rata nilai *silhouette*

Dalam penelitian [19], semakin nilai rata-rata *Silhouette Coefficient* mendekati nilai 1, maka semakin baik pengelompokan data di dalam satu *cluster*. Sebaliknya jika nilai rata-rata *Silhouette Coefficient* mendekati -1, maka semakin buruk pengelompokan data di dalam satu *cluster*. Berikut kriteria pengukuran nilai *Silhouette Coefficient* [Rousseeuw, 1987] dalam [19] diuraikan dalam tabel 1.

TABEL 1.  
UKURAN NILAI SILHOUETTE COEFFICIENT

SILHOUETTE COEFFICIENT	INTERPRETASI YANG DIUSULKAN
$0.7 < SC \leq 1.0$	STRONG STRUCTURE
$0.5 < SC \leq 0.7$	MEDIUM STRUCTURE
$0.25 < SC \leq 0.5$	WEAK STRUCTURE
$SC \leq 0.25$	NO STRUCTURE

### III. HASIL DAN PEMBAHASAN

#### A. Analisis Perhitungan Algoritma K-Means

Dataset yang telah dikumpulkan akan dilakukan analisis menggunakan algoritma *K-Means*. Analisis yang dilakukan menggunakan empat atribut yaitu luas tanaman, luas panen, produksi, dan tambah tanam mulai tahun 2018 sampai dengan 2020. Atribut tersebut diinisialkan pada tabel 2.

TABEL 2.  
INISIAL ATRIBUT

ATRIBUT	INISIAL
LUAS TANAMAN	V1
LUAS PANEN	V2
PRODUKSI	V3
TAMBAH TANAM	V4

Luas tanaman diinisialkan sebagai V1, luas panen sebagai V2, produksi sebagai V3 dan tambah tanam sebagai V4. Penginisialan ini berguna untuk mempersingkat nama atribut yang digunakan. Kemudian pada perhitungan menggunakan Algoritma *K-Means* perlu untuk menentukan jumlah klaster yang akan digunakan dapat dilihat pada tabel 3.

TABEL 3.  
KLAUSTER POTENSI

KLAUSTER	LABEL POTENSI
C1	RENDAH
C2	SEDANG
C3	TINGGI

Adapun titik *centroid* awal ditentukan berdasarkan rata-rata pada setiap kuartil data. Penentuan centroid awal dapat dilihat pada tabel 4.

TABEL 4.  
PENENTUAN *CENTROID* AWAL

KLAUSTER	<i>CENTROID</i> AWAL
C1	RATA - RATA dari KUARTIL 1
C2	RATA - RATA dari KUARTIL 2
C3	RATA - RATA dari KUARTIL 3

Berikut data cabai besar yang disusun per kecamatan di setiap tahun agar membentuk sebuah baris dan kolom pada tabel 5:

TABEL 5.  
DATA CABAI BESAR

NO	KECAMATAN	TAHUN	V1	V2	V3	V4
1	BATAHAN	2018	23	23	937	17
2	BATAHAN	2019	16	3	118	16
3	BATAHAN	2020	21	15	570	8
4	SINUNUKAN	2018	18	12	468	9
5	SINUNUKAN	2019	9	4	156	3
6	SINUNUKAN	2020	5	5	190	0
7	BATANG NATAL	2018	17	13	527	11
8	BATANG NATAL	2019	22	9	352	18
9	BATANG NATAL	2020	14	9	306	1
...	....	....	....	....	....	....
67	NAGA JUANG	2018	16	14	546	2
68	NAGA JUANG	2019	16	1	56	14
69	NAGA JUANG	2020	21	18	685	6

Adapun langkah-langkah yang dilakukan dalam menyelesaikan perhitungan Algoritma *K-Means* adalah :

1. Menentukan Jumlah Kluster

Jumlah *Cluster* yang digunakan pada data akta kelahiran sebanyak 3 *Cluster*. *Cluster* tersebut diantaranya Potensi Rendah (C1), Sedang (C2) dan Tinggi (C3) berdasarkan data Hortikultura Sayuran Cabai Besar dari BPS pada tahun 2018, 2019 dan 2020 dengan jumlah 23 Kecamatan.

2. Menentukan *Centroid*

Menentukan pusat awal *Cluster* (*Centroid*) ditentukan secara manual yang diambil dari data yang ada dalam *range*. Nilai *Cluster* 1 diambil dari data rata-rata nilai kuartil satu, nilai *Cluster* 2 diambil dari data rata-rata kuartil 2 dan Nilai *Cluster* 3 diambil dari data rata-rata kuartil 3. *Centroid* awal pada cabai besar dapat dilihat pada tabel 6.

TABEL 6.  
*CENTROID* AWAL CABAI BESAR

LABEL POTENSI	V1	V2	V3	V4
C1	9	4	155	2
C2	15	9	338	6
C3	38	26	998	20

3. Hitung Jarak *Euclidean*

$$D(x_1, y_1) = \sqrt{(23 - 9)^2 + (23 - 4)^2 + (937 - 155)^2 + (17 - 2)^2} = 782,8927$$

$$D(x_2, y_2) = \sqrt{(23 - 15)^2 + (23 - 9)^2 + (937 - 338)^2 + (17 - 6)^2} = 598,9735$$

$$D(x_3, y_3) = \sqrt{(23 - 38)^2 + (23 - 26)^2 + (937 - 998)^2 + (17 - 20)^2} = 62,76401$$

.....

Dilanjutkan sampai pada semua baris data terakhir dengan cara menghitung jarak euclidean pada variabel v1, v2,v3 dan v4 dengan titik pusat awal yang telah ditentukan.

4. Memilih nilai terkecil dari semua jarak klaster yang telah dihitung. Berikut diuraikan dalam bentuk tabel 7.

TABEL 7.  
JARAK EUCLIDEAN ITERASI-1 CABAI BESAR

NO	C1	C2	C3	JARAK TERKECIL	LABEL POTENSI
1	782,8927	598,9735	62,76401	62,76401	C3
2	39,83368	220,6443	880,444	39,83368	C1
3	415,7559	231,8181	428,5017	231,8181	C2
4	313,703	129,757	530,5316	129,757	C2
5	1,714499	182,5407	842,8144	1,714499	C1
6	35,68315	148,8716	809,0481	35,68315	C1
7	372,6966	188,7742	471,5868	188,7742	C2
8	198,5319	19,33492	646,2845	19,33492	C2
9	151,5625	32,77148	692,7416	32,77148	C2
...	....	....	....	....	....
...	....	....	....	....	....
67	391,5835	207,7625	452,904	207,7625	C2
68	99,62945	282,5633	942,4674	99,62945	C1
69	530,7289	346,8257	313,7264	313,7264	C3

#### 5. Hitung Centroid Baru

Setelah diketahui hasil kelompok pada iterasi sebelumnya kemudian pusat centroid baru dihitung berdasarkan data tiap-tiap klaster sesuai dengan rumus pada persamaan (2).

Centroid baru C1 dihitung berdasarkan jumlah seluruh data pada suatu variabel dengan hasil klaster yang sama yaitu C1 kemudian dibagi dengan banyaknya anggota pada suatu klaster itu.

$$V1 : (16 + 9 + 5 + \dots + \dots + 16) / 29 = 13,0690$$

$$V2 : (3 + 4 + 5 + \dots + \dots + 1) / 29 = 4,4138$$

$$V3 : (118 + 156 + 190 + \dots + \dots + 56) / 29 = 170,7586$$

$$V4 : (16 + 3 + 0 + \dots + \dots + 14) / 29 = 6,8621$$

Centroid baru C2 dihitung berdasarkan jumlah seluruh data pada suatu variabel dengan hasil klaster yang sama yaitu C2 kemudian dibagi dengan banyaknya anggota pada suatu klaster itu.

$$V1 : (21 + 18 + 17 + 22 + 14 + \dots + \dots + 16) / 28 = 16,5714$$

$$V2 : (15 + 12 + 13 + 9 + 9 + \dots + \dots + 14) / 28 = 11,3571$$

$$V3 : (570 + 468 + 527 + 352 + 306 + \dots + \dots + 546) / 28 = 438,7500$$

$$V4 : (8 + 9 + 11 + 18 + 1 + \dots + \dots + 2) / 28 = 6,7143$$

Centroid baru C3 dihitung berdasarkan jumlah seluruh data pada suatu variabel dengan hasil klaster yang sama yaitu C3 kemudian dibagi dengan banyaknya anggota pada suatu klaster itu.

$$V1 : (23 + \dots + \dots + 21) / 12 = 48,5000$$

$$V2 : (23 + \dots + \dots + 18) / 12 = 36,5833$$

$$V3 : (937 + \dots + \dots + 685) / 12 = 1421,0000$$

$$V4 : (17 + \dots + \dots + 6) / 12 = 21,5000$$

Hasil perhitungan centroid baru disajikan kedalam tabel 8.

TABEL 8.  
CENTROID BARU ITERASI-1 CABAI BESAR

KLASTER	V1	V2	V3	V4
C1	13,0690	4,4138	170,7586	6,8621
C2	16,5714	11,3571	438,7500	6,7143
C3	48,5000	36,5833	1421,0000	21,5000

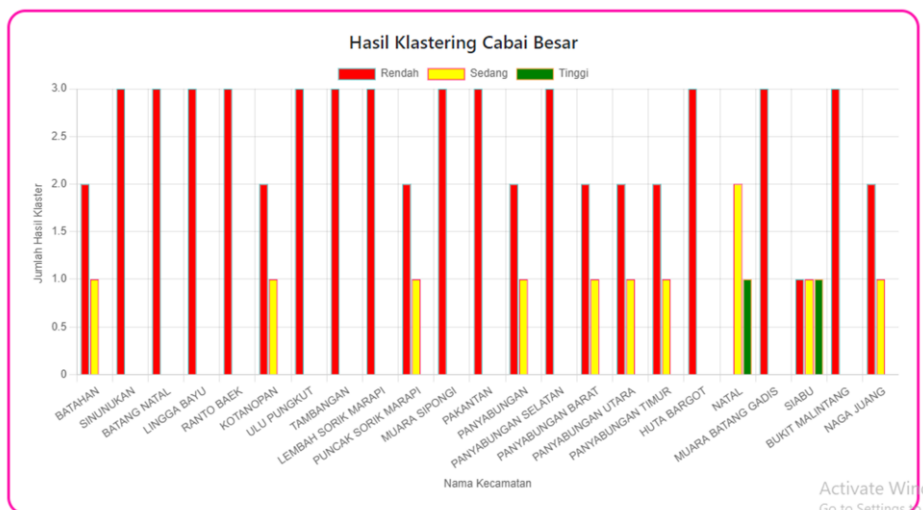
Langkah selanjutnya adalah menghitung *Euclidean Distance* kembali menggunakan titik pusat centroid baru. Jika hasil kluster pada iterasi pertama tidak berubah maka perhitungan dihentikan namun apabila kluster berubah maka perhitungan jarak dilanjutkan. Berikut hasil akhir klustering pada tabel 9.

TABEL 9.  
HASIL AKHIR KLASTERING CABAI BESAR

NO	KECAMATAN	TAHUN	V1	V2	V3	V4	POTENSI
1	BATAHAN	2018	23	23	937	17	C2 (SEDANG)
2	BATAHAN	2019	16	3	118	16	C1 (RENDAH)
3	BATAHAN	2020	21	15	570	8	C1 (RENDAH)
4	SINUNUKAN	2018	18	12	468	9	C1 (RENDAH)
5	SINUNUKAN	2019	9	4	156	3	C1 (RENDAH)
6	SINUNUKAN	2020	5	5	190	0	C1 (RENDAH)
7	BATANG NATAL	2018	17	13	527	11	C1 (RENDAH)
8	BATANG NATAL	2019	22	9	352	18	C1 (RENDAH)
9	BATANG NATAL	2020	14	9	306	1	C1 (RENDAH)
...	....	....	....	....	....	....	....
67	NAGA JUANG	2018	16	14	546	2	C1 (RENDAH)
68	NAGA JUANG	2019	16	1	56	14	C1 (RENDAH)
69	NAGA JUANG	2020	21	18	685	6	C2 (SEDANG)

B. Hasil Akhir Klustering Cabai Besar

Berikut hasil klustering cabai besar divisualisasikan menggunakan grafik pada gambar 2.



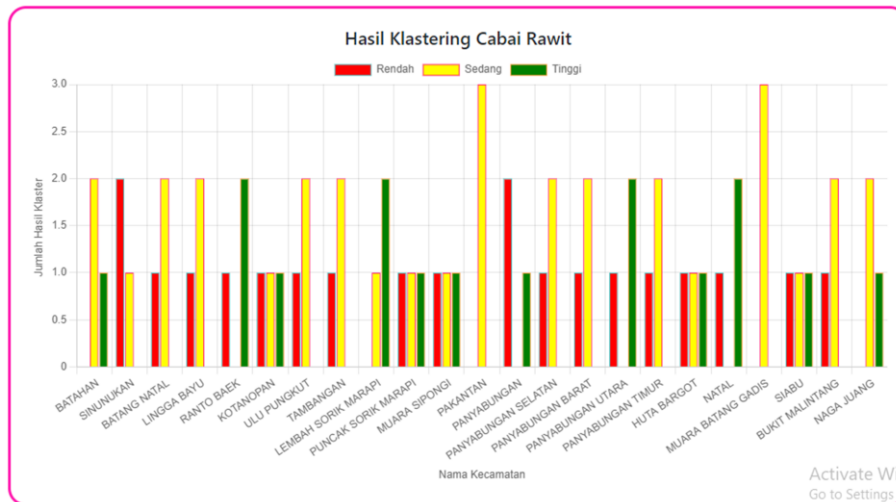
Gambar 2. Grafik kluster cabai besar

Hasil klustering Cabai Besar berhenti pada iterasi ke-12 dengan jumlah C1 (Rendah) sebanyak 56, C2 (Sedang) sebanyak 11 dan C3 (Tinggi) sebanyak 2. Adapun hasil analisa klustering Cabai Besar yang menjadi potensi unggul di setiap kecamatan adalah kecamatan Siabu tahun 2018 dan Natal 2020 yang termasuk kedalam kluster C3 (Tinggi).

C. Hasil Akhir Klustering Cabai Rawit

Berikut hasil klustering cabai rawit divisualisasikan menggunakan grafik pada gambar 3.



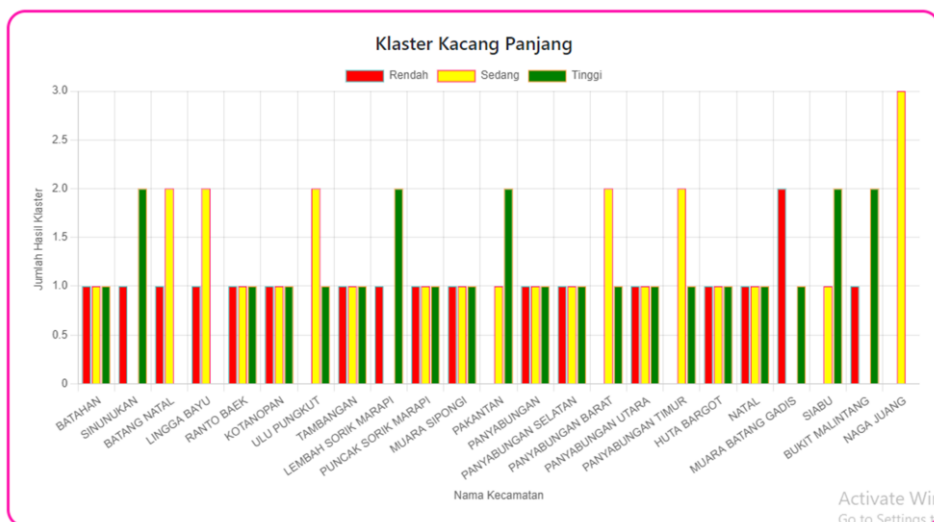


Gambar 3. Grafik hasil klaster cabai rawit

Hasil klastering Cabai Rawit berhenti pada iterasi ke-2 dengan jumlah C1 (Rendah) sebanyak 20, C2 (Sedang) sebanyak 33 dan C3 (Tinggi) sebanyak 16. Adapun hasil analisa klastering Cabai Rawit yang menjadi potensi unggul di setiap kecamatan adalah kecamatan Batahan tahun 2018, Ranto Baik tahun 2018 dan 2020, Kotanopan tahun 2018, Lembah Sorik Marapi tahun 2018 dan tahun 2020, Puncak Sorik Marapi tahun 2018, Muara Sipongi tahun 2020, Panyabungan tahun 2019, Panyabungan Utara tahun 2018 dan 2020, Huta Bargot tahun 2020, Natal tahun 2018 dan 2020, Siabu tahun 2018 dan Naga Juang tahun 2018.

**D. Hasil Akhir Klastering Kacang Panjang**

Berikut hasil klastering kacang panjang divisualisasikan menggunakan grafik pada gambar 4.



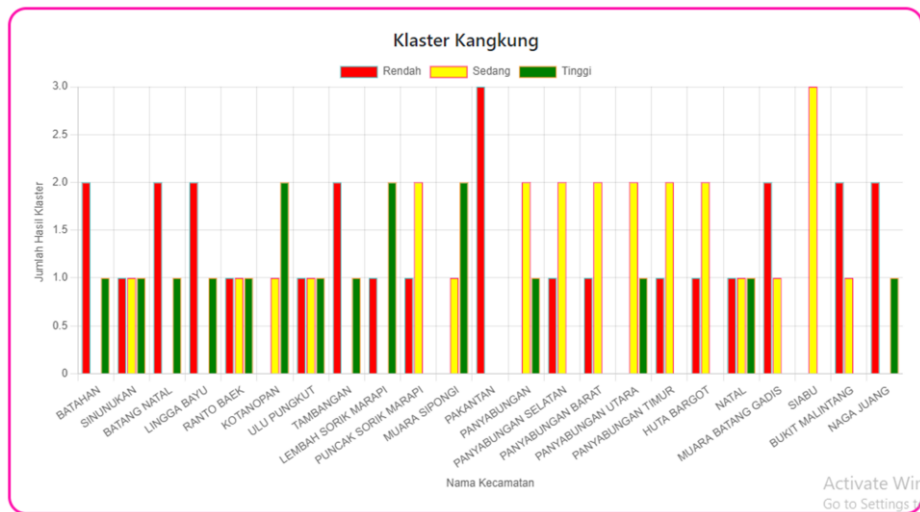
Gambar 4. Grafik hasil klaster kacang panjang

Hasil klastering Kacang Panjang berhenti pada iterasi ke-2 dengan jumlah C1 (Rendah) sebanyak 18, C2 (Sedang) sebanyak 26 dan C3 (Tinggi) sebanyak 25. Adapun hasil analisa klastering Kacang Panjang yang menjadi potensi sangat unggul di setiap kecamatan adalah Batahan tahun 2020, Sinunukan tahun 2019 dan 2020, Ranto Baik tahun 2019, Kotanopan, Ulu Pungkut, dan Tambangan pada tahun 2020, Lembah Sorik Marapi tahun 2019 dan 2020, Puncak Sorik Marapi dan Muara Sipongi pada tahun 2020, Pakantan tahun 2019 dan 2020, Panyabungan, Panyabungan Selatan,

Panyabungan Utara, Panyabungan Timur, Panyabungan Barat, Huta Bargot, Natal, dan Muara Batang Gadis pada tahun 2020, Siabu tahun 2018 dan 2020, Bukit Malintang tahun 2018 dan 2020.

**E. Hasil Akhir Klastering Kangkung**

Berikut hasil klastering kangkung divisualisasikan menggunakan grafik pada gambar 5.

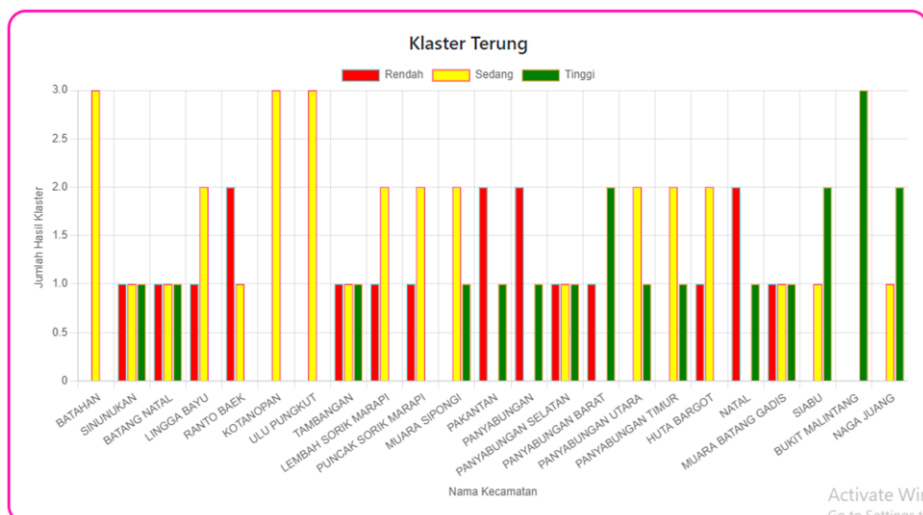


Gambar 5. Grafik hasil klaster kangkung

Hasil klastering kangkung berhenti pada iterasi ke-3 dengan jumlah C1 (Rendah) sebanyak 27, C2 (Sedang) sebanyak 25 dan C3 (Tinggi) sebanyak 17. Adapun hasil analisa klastering kangkung yang menjadi potensi unggul di setiap kecamatan adalah Batahan, Sinunukan, Batang Natal, Lingga Bayu, Ranto Baek, Ulu Pungkut, Tambangan pada tahun 2020, Kotanopan tahun 2018 dan 2020, Lembah Sorik Marapi tahun 2019 dan 2020, Muara Sipongi tahun 2019 dan 2020, Panyabungan tahun 2018, Panyabungan Utara, Natal dan Naga Juang pada tahun 2020.

**F. Hasil Akhir Klastering Terung**

Berikut hasil klastering terung divisualisasikan menggunakan grafik pada gambar 6.

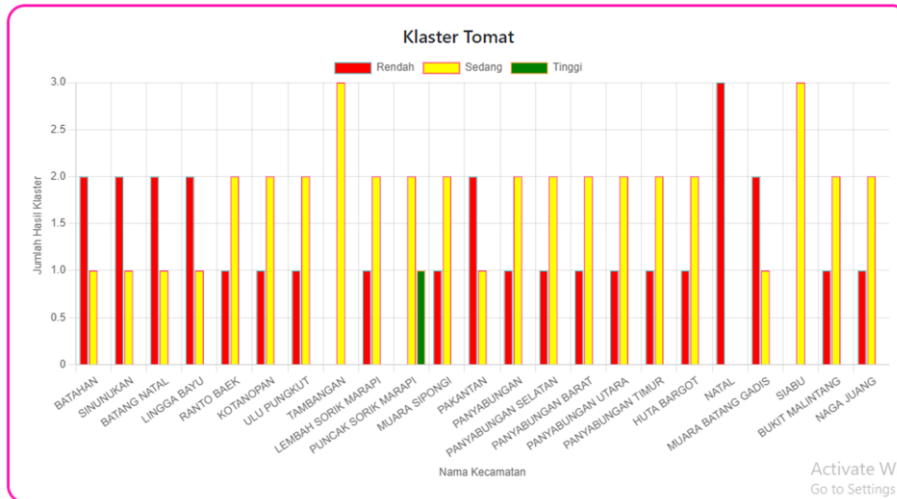


Gambar 6. Grafik klaster terung

Hasil klastering Terung berhenti pada iterasi ke-3 dengan jumlah C1 (Rendah) sebanyak 30, C2 (Sedang) sebanyak 20 dan C3 (Tinggi) sebanyak 19. Adapun hasil analisa klastering Terung yang menjadi potensi unggul di setiap kecamatan adalah Sinunukan tahun 2020, Batang Natal tahun 2018, Tambangan tahun 2019, Muara Sipongi tahun 2018, Pakantan, Panyabungan dan Panyabungan Selatan pada tahun 2020, Panyabungan Barat tahun 2018 dan 2020, Panyabungan Utara tahun 2018, Panyabungan Timur tahun 2020, Natal tahun 2020, Muara Batang Gadis tahun 2018, Siabu tahun 2018 dan 2020, Bukit Malintang tahun 2018,2019, dan 2020, Naga Juang tahun 2018 dan 2020.

G. Hasil Akhir Klastering Tomat

Berikut hasil klastering tomat divisualisasikan menggunakan grafik pada gambar 7.



Gambar 7. Grafik klaster tomat

Hasil klastering Tomat berhenti pada iterasi ke-6 dengan jumlah C1 (Rendah) sebanyak 28, C2 (Sedang) sebanyak 40 dan C3 (Tinggi) sebanyak 1. Adapun hasil analisa klastering Tomat yang menjadi potensi unggul di kecamatan adalah kecamatan Puncak Sorik Marapi tahun 2018 yang termasuk kedalam klaster C3 (Tinggi).

H. Implementasi Sistem

Berikut gambar 8 hasil implementasi sistem yang telah dibuat.

NO.	KECAMATAN	TAHUN	LUAS TANAMAN	LUAS PANEN	PRODUKSI	TAMBAH TANAM	AKSI
1.	BATAHAN	2018	23	23	937	17	[Add] [Edit] [Delete]
2.	BATAHAN	2019	16	3	118	16	[Add] [Edit] [Delete]
3.	BATAHAN	2020	21	15	570	8	[Add] [Edit] [Delete]
4.	SINUNUKAN	2018	18	12	468	9	[Add] [Edit] [Delete]
5.	SINUNUKAN	2019	9	4	156	3	[Add] [Edit] [Delete]
6.	SINUNUKAN	2020	5	5	190	0	[Add] [Edit] [Delete]
7.	BATANG NATAL	2018	17	13	527	11	[Add] [Edit] [Delete]
8.	BATANG NATAL	2019	22	9	352	18	[Add] [Edit] [Delete]
9.	BATANG NATAL	2020	14	9	306	1	[Add] [Edit] [Delete]
10.	LINGGA BAYU	2018	18	11	430	8	[Add] [Edit] [Delete]

Gambar 8. Halaman crud data

Pada gambar 8 halaman CRUD data, sistem dapat menampilkan data yang sudah di input kedalam database. Pada halaman ini user juga dapat untuk menambah, mengedit dan menghapus data ke dalam database. Kemudian proses klastering dapat dilihat pada gambar 9.



#### IV. SIMPULAN

Klastering sayuran unggulan menggunakan algoritma *K-Means* telah berhasil dilakukan yang dikelompokkan menjadi 3 kluster yaitu kluster rendah, sedang, dan tinggi serta didapatkan output berupa sistem berbasis web dalam pengaplikasiannya. Sayuran dengan jumlah C3 (Tinggi) terbanyak pada setiap kecamatan adalah Kacang Panjang kemudian setelahnya yaitu Terung, Kangkung, Cabai Rawit, Cabai Besar dan Tomat. Hasil evaluasi klastering menggunakan *silhouette* adalah baik (medium structure) dengan *score* = 0,6798 yang menggunakan jumlah  $k = 3$  kluster. Pada penelitian selanjutnya dapat mencoba menggunakan jumlah  $k$  terbaik untuk mendapatkan kualitas hasil klastering yang paling baik.

#### DAFTAR PUSTAKA

- [1] BPS, *Analisis Komoditas Unggulan Hortikultura Sayur-Sayuran Provinsi Sumatera Utara 2015-2019*. Medan, 2021.
- [2] E. Humaidi, B. Unteawati, and A. Analianasari, "Pemetaan Komoditas Sayur Unggulan Di Provinsi Lampung," *J. Agribisnis Indones.*, vol. 8, no. 2, pp. 106–114, 2020.
- [3] BPS, *Statistik Tanaman Hortikultura Provinsi Sumatera Utara 2020*. Medan: Badan Pusat Statistik Provinsi Sumatera Utara, 2021.
- [4] M. Sadli, F. Fajriana, W. Fuadi, E. Ermatita, and I. Pahendra, "Penerapan Model K-Nearest Neighbors Dalam Klasifikasi Kebutuhan Daya Listrik Untuk Masing-Masing Daerah Di Kota Lhokseumawe," *J. ECOTIPE*, vol. 5, no. 2, pp. 11–18, 2018.
- [5] S. Wijayanto and Y. Fathoni, M, "Pengelompokan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means," *J. JUPITER*, vol. 13, no. 2, pp. 212–219, 2021.
- [6] W. Pulungan, P. Poningsih, and H. Satria, "Pengelompokan Pada Kendaraan Bermotor Menurut Kegunaannya Menggunakan Metode Data Mining K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 746–752, 2019.
- [7] Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 192, 2020.
- [8] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, p. 167, 2018.
- [9] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 6, no. 2, pp. 229–240, 2021.
- [10] R. D. Kusumah, B. Warsito, and M. A. Mukid, "Perbandingan Metode K-Means Dan Self Organizing Map (Studi Kasus: Pengelompokan Kabupaten/Kota Di Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia 2015)," *J. Gaussian*, vol. 6, no. 3, pp. 429–437, 2017.
- [11] Y. W. Harinta and S. S. Joko Setyo Basuki, "Pemetaan dan Pengembangan Agribisnis Komoditas Unggulan Sayuran di Kabupaten Karanganyar," *J. Sos. Ekon. dan Kebijakan. Pertan.*, vol. 7, no. 2, 2018.
- [12] M. W. Talakua, Z. A. Leleury, and A. W. Talluta, "Acluster Analysis By Using K-Means Method for Grouping of District/City in Maluku Province Industrial Based on Indicators of Maluku Development Index in 2014," *Barekeng J. Ilmu Mat. dan Terap.*, vol. 11, no. 2, pp. 119–128, 2017.
- [13] 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.
- [14] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data," *Qual Quant*, vol. 56, pp. 1283–1291, 2022.
- [15] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017.
- [16] A. S. Ahmar, D. Napitupulu, R. Rahim, R. Hidayat, Y. Sonatha, and M. Azmi, "Using K-Means Clustering to Cluster Provinces in Indonesia," *J. Phys. Conf. Ser.*, vol. 1028, no. 1, 2018.
- [17] D. Widyadhana, R. B. Hastuti, I. Kharisudin, and F. Fauzi, "Perbandingan Analisis Kluster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah," *Prism. Pros. Semin. Nas. Mat.* 4, vol. 4, pp. 584–594, 2021.
- [18] W. Yustanti, N. Rahmawati, and Y. Yamasari, "Klastering Wilayah Kota/Kabupaten Berdasarkan Data Persebaran Covid-19 Di Propinsi Jawa Timur dengan Metode K-Means," *JIEET (Journal Inf. Eng. Educ. Technol.)*, vol. 4, no. 1, pp. 1–9, 2020.
- [19] 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.
- [20] S. Paembonan and H. Abdul, "Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat," *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, p. 48, 2021.