



## Penentuan Daerah Sentra Produksi Jagung (*Zea mays L.*) di Kabupaten/Kota Provinsi Sumatera Barat dengan Metode K-Means Clustering

Santosa<sup>1\*</sup>, Andasuryani<sup>2</sup>, Saddam Pebrianto<sup>2</sup>, Raihan M. Deri Prasmadika<sup>2</sup>

<sup>1</sup> Departemen Teknologi Industri Pertanian, Fakultas Teknologi Pertanian, Universitas Andalas, Indonesia

<sup>2</sup> Departemen Teknik Pertanian dan Biosistem, Fakultas Teknologi Pertanian, Universitas Andalas, Indonesia

### ABSTRAK

Jagung (*Zea mays L.*) merupakan komoditas pangan strategis yang berperan penting dalam ketahanan pangan dan industri. Provinsi Sumatera Barat memiliki potensi besar dalam pengembangan komoditas ini, namun persebaran produksi yang tidak merata memerlukan identifikasi wilayah sentra produksi yang tepat. Penelitian ini bertujuan untuk menentukan daerah-daerah potensial sebagai sentra produksi jagung dengan menerapkan metode *K-Means Clustering*. Data yang digunakan merupakan data sekunder tahun 2017–2022 dari 19 kabupaten/kota, mencakup lima variabel: luas panen, produksi, produktivitas, ketinggian wilayah, dan curah hujan. Metode Elbow digunakan untuk menentukan jumlah klaster optimal. Hasil analisis menunjukkan bahwa beberapa daerah secara konsisten berada dalam klaster dengan nilai di atas rata-rata, antara lain Kabupaten Pasaman Barat, Agam, Pesisir Selatan, Pasaman, dan Solok Selatan. Penelitian ini memberikan informasi berbasis data yang berguna dalam perencanaan pengembangan jagung secara spasial di Provinsi Sumatera Barat.

### KATA KUNCI

Jagung; K-Means Clustering; Data Mining; Sentra Produksi; Sumatera Barat.

### PENULIS KORESPONDEN

Alamat e-mail penulis koresponden: [santosa@ae.unand.ac.id](mailto:santosa@ae.unand.ac.id)

### 1. Pendahuluan

Jagung merupakan salah satu tanaman pangan utama dunia, sejajar dengan padi dan gandum, dan berperan sebagai sumber pangan alternatif yang signifikan di Amerika Serikat, Amerika Latin, serta berbagai negara Asia, termasuk Indonesia. Jagung di Indonesia tidak hanya dikonsumsi sebagai bahan pangan pokok kedua setelah padi, tetapi juga digunakan sebagai bahan baku industri dan pakan ternak [1], [2].

Berdasarkan data Badan Pusat Statistik (BPS) Sumatera Barat, produksi jagung pada tahun 2020 mencapai 939.465,95 ton dengan luas panen sebesar 134.911,70 hektar. Pada tahun 2021, terjadi peningkatan produksi menjadi 948.063,16 ton dengan luas panen 134.671,20 hektar [3]. Namun demikian, distribusi produksi jagung di Sumatera Barat masih belum merata, sehingga diperlukan suatu pendekatan ilmiah untuk mengidentifikasi wilayah-wilayah yang layak dikembangkan sebagai sentra produksi. Untuk tujuan tersebut, pemanfaatan metode pengelompokan data (*Clustering*) menjadi sangat relevan. Salah satu algoritma yang umum digunakan dalam *data mining* adalah metode *K-means Clustering*, yang memiliki kemampuan untuk mengelompokkan data secara objektif berdasarkan kemiripan karakteristik antar wilayah tanpa memerlukan variabel target [4].

Penelitian ini menggunakan lima variabel utama, yaitu (1) Luas panen (hektar) : menunjukkan skala budidaya jagung; (2) Produksi (ton) : merepresentasikan volume hasil panen; (3) Produktivitas (kuintal/hektar) : mengukur efisiensi hasil per satuan luas; (4) Tinggi wilayah (mdpl) : karena ketinggian berpengaruh terhadap suhu dan jenis tanah; (5) Curah hujan (mm/tahun) : sebagai faktor penting dalam ketersediaan air dan pertumbuhan tanaman jagung. Dengan menggunakan pendekatan K-means Clustering pada kelima variabel tersebut, diharapkan hasilnya dapat mendukung pengambilan kebijakan yang lebih tepat sasaran dalam pengembangan sentra produksi jagung di Sumatera Barat.



Jagung memiliki karakter adaptif yang relatif luas terhadap berbagai kondisi agroekologi, sehingga komoditas ini kerap dijadikan penopang ketahanan pangan dan ekonomi pertanian di banyak wilayah. Selain perannya dalam memenuhi kebutuhan konsumsi langsung, jagung juga menjadi bahan strategis dalam rantai pasok industri pangan olahan, bioenergi, serta pakan ternak, yang permintaannya cenderung meningkat seiring pertumbuhan penduduk dan sektor peternakan. Kondisi tersebut menjadikan jagung tidak hanya bernilai agronomis, tetapi juga bernilai ekonomis dan strategis dalam perencanaan pembangunan pertanian daerah [5], [6].

Di tingkat regional, Sumatera Barat memiliki keragaman topografi dan iklim yang cukup kompleks, mulai dari dataran rendah pesisir hingga kawasan dataran tinggi pegunungan. Keragaman ini berdampak langsung pada variasi produktivitas jagung antardaerah, baik dari segi kuantitas hasil maupun stabilitas produksi tahunan. Peningkatan produksi yang tercatat secara agregat belum tentu mencerminkan kinerja yang merata pada seluruh kabupaten dan kota, karena masih terdapat wilayah yang potensinya belum optimal akibat keterbatasan teknologi, akses sarana produksi, maupun kesesuaian lahan [7], [8].

Ketidakseimbangan distribusi produksi tersebut menuntut adanya analisis berbasis data yang mampu menggambarkan pola dan karakteristik setiap wilayah secara komprehensif. Pendekatan konvensional yang hanya mengandalkan satu indikator, seperti total produksi atau luas panen, sering kali kurang memadai untuk menangkap kompleksitas faktor-faktor yang memengaruhi keberhasilan budidaya jagung. Oleh karena itu, diperlukan metode analisis multivariat yang dapat mengintegrasikan berbagai variabel sekaligus guna memperoleh gambaran yang lebih utuh mengenai kondisi wilayah produksi [9], [10].

Metode K-means Clustering dipandang sesuai untuk tujuan tersebut karena mampu mengelompokkan wilayah berdasarkan tingkat kemiripan karakteristik tanpa asumsi awal mengenai struktur data. Hasil pengelompokan yang dihasilkan dapat menunjukkan kelompok wilayah dengan potensi tinggi, sedang, maupun rendah dalam pengembangan jagung, sehingga memudahkan penentuan prioritas intervensi. Selain itu, pendekatan ini bersifat objektif dan replikatif, sehingga dapat digunakan kembali pada periode waktu yang berbeda untuk memantau dinamika perubahan produksi [11].

Pemilihan variabel luas panen, produksi, dan produktivitas memberikan gambaran menyeluruh mengenai aspek skala usaha, hasil aktual, serta efisiensi pemanfaatan lahan. Sementara itu, variabel ketinggian wilayah dan curah hujan merepresentasikan faktor lingkungan yang secara ilmiah terbukti memengaruhi pertumbuhan dan perkembangan tanaman jagung. Integrasi antara variabel agronomis dan agroklimat ini memungkinkan analisis yang tidak hanya bersifat deskriptif, tetapi juga interpretatif terhadap kondisi biofisik wilayah. Hasil pengelompokan yang diperoleh tidak sekadar berfungsi sebagai klasifikasi statistik, melainkan juga sebagai dasar perumusan kebijakan pengembangan pertanian yang berbasis potensi lokal. Pemerintah daerah dan pemangku kepentingan terkait dapat memanfaatkan informasi ini untuk menentukan strategi intensifikasi atau ekstensifikasi, alokasi bantuan sarana produksi, pengembangan infrastruktur pertanian, serta perencanaan penyuluhan yang lebih terarah. Pada akhirnya, pendekatan ini diharapkan mampu mendorong peningkatan produktivitas jagung secara berkelanjutan sekaligus mengurangi kesenjangan antardaerah di Sumatera Barat.

## 2. Metode Penelitian

### 2.1 Waktu dan Tempat

Penelitian ini dilaksanakan pada bulan Oktober 2024 - Februari 2025 di Laboratorium Agricultural Management and Geographic Information System (AMGIS) Departemen Teknik Pertanian dan Biosistem, Fakultas Teknologi Pertanian, Universitas Andalas.



## 2.2 Alat dan Bahan

Alat yang akan digunakan pada penelitian ini adalah Laptop, aplikasi Microsoft Excel. Bahan yang akan digunakan pada penelitian ini adalah Data sekunder yang didapatkan dari Badan Pusat Statistik (BPS) pada tahun 2017 hingga tahun 2022 berupa Luas Panen (hektar), Produksi (ton), Produktivitas (kuintal/hektar), Tinggi Wilayah (mdpl), Curah Hujan (mm/Tahun) di setiap Kabupaten/Kota di Sumatera Barat.

## 2.3 Metode Penelitian

### 2.3.1 Identifikasi Variabel

Data yang dipakai dalam penelitian berasal dari database Badan Pusat Statistik (BPS) Sumatera Barat. Variabel dalam penelitian ini mencakup pengetahuan tentang pengambilan keputusan daerah potensi tanaman jagung, yang didasarkan pada studi literatur dan rentang waktu yang ditentukan.

### 2.3.2 Pengumpulan Data Series

Data yang terkumpul untuk penelitian ini adalah rangkaian variabel yang terkait dengan periode waktu 2017-2022. Database tersebut mencakup luas panen (hektar), produksi (ton), produktivitas (kuintal/hektar), Tinggi Wilayah (mdpl), Curah Hujan (mm/tahun) di setiap kabupaten dan kota di Provinsi Sumatera Barat.

### 2.3.3 Data Pre-processing

Data series tahunan yang terkumpul kemudian diekstraksi (baik dalam bentuk tabel maupun data *software*) menjadi *database* yang sudah diintegrasikan untuk memudahkan pengolahan data. Tahap ini melibatkan pengolahan data yang didapatkan dari Badan Pusat Statistik Provinsi Sumatera Barat, yang kemudian diimport ke dalam lembar kerja Microsoft Excel. Beberapa proses pengolahan data yang dilakukan meliputi:

#### a) Data Cleaning

Proses data cleaning adalah langkah untuk menghapus data yang tidak lengkap dan tidak konsisten. Data yang memiliki nilai kosong dan tidak konsisten akan diproses *cleaning* dengan cara menghapusnya agar tidak mengganggu proses *input* pada Clustering *K-means*. Data *integration*.

#### b) Data Integration

Data *Integration* adalah proses penggabungan data dari berbagai sumber. Data series tahun 2017-2022 memiliki struktur yang sama, tetapi tersimpan dalam *file Excel* yang berbeda-beda. Proses integrasi ini bertujuan untuk menggabungkan data dari beberapa sumber ke dalam penyimpanan data sesuai format yang telah ditentukan.

#### c) Data Selection

Data *selection* adalah proses penting dalam *K-means Clustering* di mana atribut-atribut data yang akan digunakan sebagai dataset dipilih. Tidak semua data yang tersedia dalam database digunakan, sehingga hanya data yang relevan dan sesuai yang diambil untuk dianalisis. Tujuan dari proses ini adalah untuk memastikan bahwa *dataset* yang digunakan dalam proses *K-means Clustering* hanya terdiri dari data yang memiliki nilai informasi yang tinggi dan dapat memberikan kontribusi yang signifikan dalam menghasilkan klaster yang akurat dan bermakna. Dengan memilih data yang sesuai, proses *Clustering* dapat dilakukan dengan lebih efisien dan menghasilkan hasil yang lebih relevan dan bermakna.

#### d) Data Transformation

Data *transformation* adalah tahap penting dalam analisis klasterisasi dengan *K-means Clustering*, di mana data awal diubah menjadi bentuk yang lebih sesuai untuk pengolahan dengan metode *K-means Clustering* ini. Tujuan dari data *transformation* adalah untuk memastikan bahwa data yang digunakan memenuhi asumsi dan persyaratan *K-means Clustering*. Dengan data yang tepat, kita dapat memastikan bahwa data yang digunakan untuk *K-means Clustering* memenuhi asumsi dasar algoritma tersebut. Sebelum data dianalisis dengan metode *K-Means Clustering*, diperlukan tahap

transformasi berupa standardisasi data. Hal ini penting dilakukan karena setiap variabel dalam penelitian ini memiliki satuan dan skala yang berbeda, misalnya luas panen dalam hektar, produksi dalam ton, produktivitas dalam kuintal/hektar, ketinggian wilayah dalam meter, dan curah hujan dalam milimeter/tahun. Perbedaan skala ini dapat menyebabkan dominasi salah satu variabel terhadap hasil pengelompokan (*clustering*) jika tidak dinormalisasi terlebih dahulu. Salah satu metode yang digunakan untuk menstandarisasi data adalah *Z-Score Standardization*. Metode ini mengubah setiap nilai dalam dataset menjadi bentuk yang memiliki rata-rata nol dan standar deviasi satu, sehingga semua variabel memiliki skala yang sebanding. *Z-Score* dihitung menggunakan rumus sebagai berikut:

## Keterangan :

Z = nilai hasil standardisasi

X = nilai asli data

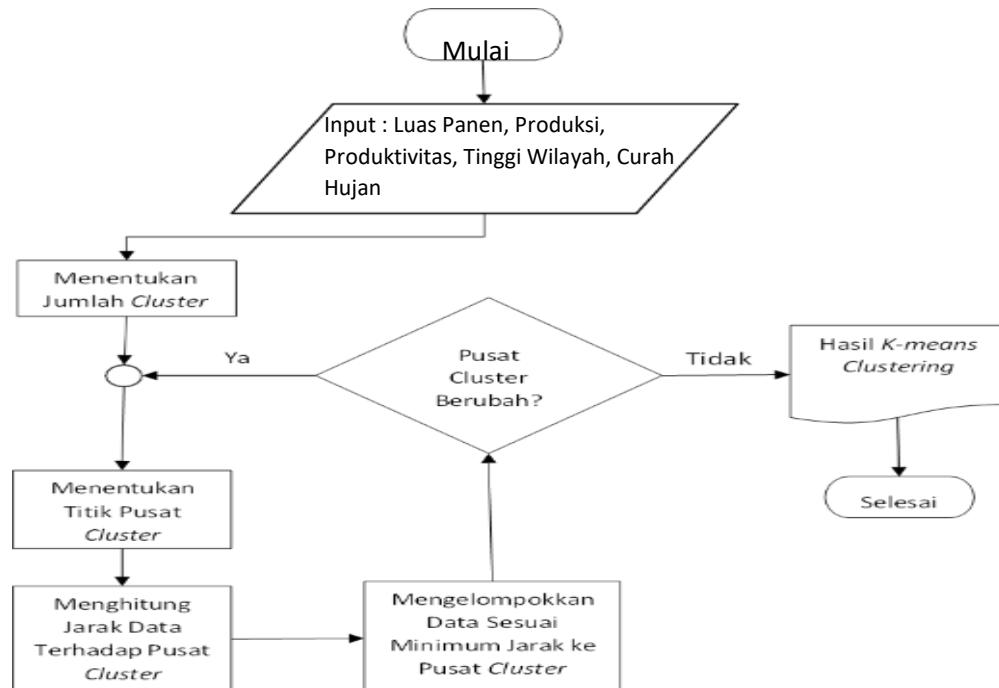
$\mu$  = rata-rata dari data

$\sigma$  = standar deviasi dari data

Melalui proses ini, nilai-nilai yang memiliki satuan dan skala berbeda dapat dibandingkan secara proporsional. Data yang telah dinormalisasi menggunakan Z-Score kemudian digunakan dalam proses pengelompokan dengan algoritma K-Means untuk menghasilkan hasil yang lebih objektif dan valid [12], [13].

#### **2.3.4 Proses *K-means Clustering***

Data yang digunakan dalam penelitian ini merupakan data series yang diperoleh dari Badan Pusat Statistik (BPS), yang mencakup periode tahun 2017 hingga 2022. Adapun beberapa langkah yang perlu dilakukan dalam melakukan Clustering dengan menggunakan algoritma K-means, dapat dilihat melalui flow chart pada Gambar 1.



**Gambar 1. Flowchart Proses K-Means Clustering**



*K-means Clustering* adalah metode analisis data atau *data mining* tanpa pengawasan yang mengelompokkan data ke dalam kelompok berdasarkan karakteristik yang sama. Algoritma *K-means* membagi data menjadi k klaster dengan tingkat kemiripan tinggi dalam satu klaster dan rendah dengan klaster lainnya. Dari gambar di atas dapat dijelaskan langkah-langkah *K-means Clustering* sebagai berikut:

- a. Memilih jumlah klaster k. Dalam penerapan algoritma *K-Means Clustering*, jumlah klaster (k) harus ditentukan di awal sebagai parameter utama dalam proses pengelompokan. Dalam penelitian ini, digunakan rentang jumlah klaster dari 2 hingga 6. Pemilihan rentang tersebut didasarkan pada pertimbangan *metodologis* dan konteks data. Jumlah objek yang dianalisis sebanyak 19 kabupaten/kota, sehingga jumlah klaster yang ideal secara statistik umumnya berkisar antara 2 hingga akar dari jumlah objek. Jumlah klaster yang optimal secara matematis dapat diperkirakan dengan rumus akar kuadrat dari n, di mana n adalah jumlah total data. Dengan demikian, akar kuadrat dari 19 yaitu 4,35 sehingga pemilihan rentang 2 sampai 6 klaster dianggap wajar dan dapat mencerminkan variasi karakteristik wilayah tanpa menghasilkan pembagian yang terlalu sempit maupun terlalu luas [14].
- b. Menentukan titik pusat awal klaster (*centroid*). Tahapan ini dapat dilakukan dengan berbagai metode, namun yang umum dilakukan adalah menginisialisasi pusat klaster dengan angka acak. Mengalokasikan semua data/objek ke klaster terdekat. Kedekatan antara dua objek ditentukan berdasarkan jarak di antara keduanya. Demikian pula, kedekatan data dengan klaster tertentu ditentukan oleh jarak antara data dan pusat klaster. Pada tahap ini, perlu menghitung jarak setiap data dengan setiap pusat klaster. Jarak antara satu data dengan satu klaster tertentu menentukan di klaster mana data tersebut ditempatkan. Untuk menghitung jarak antara semua data dengan setiap pusat klaster, dapat menggunakan teori jarak *Euclidean*. Rumus teori *Euclidean* adalah sebagai berikut:

$$d(x, y) = \sqrt{(\sum_{k=1}^p (x_{ik} - y_{jk})^2)}$$

Keterangan :

D = jarak antara x dan y

x<sub>ik</sub> = nilai data dari titik ke-i dalam dimensi ke-k

y<sub>jk</sub> = nilai data pusat awal *cluster* dari titik ke-j dalam dimensi ke-k i = setiap data atribut

p = jarak variabel *cluster*

- c. Melakukan penghitungan ulang pusat klaster dengan anggota klaster yang sudah ada. Pusat klaster merupakan nilai rata-rata dari seluruh data/objek yang terdapat dalam klaster yang diberikan. Selain itu, opsi lain yang dapat digunakan adalah nilai median dari klaster tersebut. Dengan demikian, rata-rata bukanlah satu-satunya ukuran yang dapat digunakan dalam hal ini. Untuk menentukan pusat klaster dapat digunakan rumus berikut:

$$C = (\sum m) / n$$

Keterangan :

C = Centroid data

m = anggota data yang termasuk kedalam centroid tertentu

n = jumlah data yang menjadi anggota centroid tertentu

- d. Melakukan pengelompokan ulang terhadap setiap objek menggunakan pusat klaster yang baru. Jika pusat klaster tidak berubah, maka proses klasterisasi telah selesai. Jika pusat cluster berubah atau "Ya" seperti pada *flowchart* maka proses akan mengulang ke langkah penentuan pusat cluster baru dan kembali menghitung ulang hingga posisi pusat cluster stabil, jika "Tidak" maka proses iterasi dihentikan karena telah mencapai konvergensi. Proses selesai setelah pusat *cluster* tidak lagi berubah. Hasil akhir berupa pengelompokan daerah ke dalam beberapa *cluster* berdasarkan kesamaan karakteristik lima variabel [15].

### **2.3.5 Pengumpulan Cluster Optimum**

Setelah melakukan proses Clustering dengan menggunakan k-means, langkah berikutnya adalah menerapkan metode Elbow untuk menemukan jumlah cluster yang optimal dalam upaya mencapai hasil Clustering yang terbaik. Langkah-langkah metode elbow untuk menentukan jumlah klaster yang paling optimal adalah sebagai berikut :

1. Inisialisasi awal nilai k
  2. Menaikkan nilai k
  3. Menghitung nilai SSE (Sum of squares Error) dari setiap cluster dengan rumus:

i=1

## Keterangan :

SSE = Sum of squares Errors

$x_j$  = data pada cluster  $S_i$  (cluster 1/ cluster 2/...cluster n)

$ck$  = Centroid dari cluster atau nilai mean dari total dataset cluster

4. Melakukan perhitungan SSE (Sum of squares Error) sampai dengan k yang telah ditentukan.
  5. Melihat hasil SSE (Sum of squares Error) dari nilai yang turun secara drastis atau nilai yang berubah secara signifikan
  6. Menetapkan nilai k yang membentuk siku.

### **2.3.6 Penentuan Keputusan Sentra Produksi Jagung**

Penentuan keputusan lokasi sentra produksi jagung di Sumatera Barat dapat dibuat setelah selesai mengkalkulasi nilai rekomendasi hasil *Clustering* pada setiap variabel pada *k-means*. Hasil ini akan mengungkapkan beberapa kriteria daerah yang menduduki posisi teratas pada setiap periode tahun yang telah diolah.

### **3. Hasil dan Pembahasan**

### **3.1 Gambaran Penelitian**

Penelitian ini dilakukan dengan mengumpulkan data series terkait variabel sentra produksi jagung di Sumatera Barat. Variabel yang digunakan dalam metode *K-means* meliputi luas panen, produksi, produktivitas, ketinggian wilayah, dan curah hujan. Sebelum dianalisis, data tersebut terlebih dahulu melalui tahap *preprocessing data* untuk mengatasi perbedaan skala antar variabel serta mengurangi potensi adanya perbedaan nilai yang ekstrem, sehingga data yang digunakan lebih berkualitas dan seimbang. Proses analisis dilakukan dengan menerapkan metode *K-means*, di mana jumlah klaster per tahun telah ditetapkan sebanyak enam klaster berdasarkan data dari tahun 2017 hingga 2022. Penentuan jumlah klaster optimal dilakukan menggunakan metode *Elbow*. Selanjutnya, klaster dengan nilai pusat klaster di atas rata-rata dipilih sebagai dasar dalam merekomendasikan lokasi sentra produksi jagung di Sumatera Barat pada setiap tahunnya.

### 3.2 Identifikasi Variabel Metode K-means

### **3.2.1 Sumber Data**

Ketersediaan data dan informasi yang memadai sangat penting dalam penerapan metode *K-means Clustering*. Dalam penelitian ini, data series variabel diperoleh dalam bentuk softcopy dari instansi terkait. Data yang dikumpulkan mencakup periode 2017 hingga 2022, dengan total 570 data dari tahun 2017-2022, yang terdiri dari 19 kabupaten/kota di Sumatera Barat dan mencakup 5 variabel yang terdiri dari luas panen, produksi, produktivitas, tinggi wilayah, dan curah hujan. Sumber data yang digunakan dalam penelitian ini dapat dijelaskan sebagai berikut:



### 3.2.1.1 Data Variabel Luas Panen, Produksi, Produktivitas, dan Tinggi Wilayah

Data dalam penelitian ini diperoleh dari Badan Pusat Statistik (BPS) Sumatera Barat yang bekerja sama dengan Dinas Pertanian, Tanaman Pangan, dan Hortikultura (BPTPH) Sumatera Barat. Data yang diperoleh dapat dilihat pada Tabel 1 dan Tabel 2.

**Tabel 1. Data Luas Panen, Produksi, Produktivitas Jagung Sumatera Barat**

Kabupaten/ Kota	Luas Panen		Produksi		Produktivitas	
	2017	2018	2017	2018	2017	2018
Kab. Kepulauan Mentawai	2	5	10	25,01	50	50,02
Kab. Pesisir Selatan	20.829	18.109,1	169.102	147.080,77	81,19	81,22
Kab. Solok	397	369	4.018	3.733,26	101,21	101,17
Kab. Sijunjung	456	587,1	2.956	3.806,51	64,82	64,84
Kab. Tanah Datar	6.804	5.532,1	52.046	42.332,81	76,49	76,52
Kab. Padang Pariaman	9.400	8.992,4	80.270	76.820,3	85,39	85,43
Kab. Agam	14.444	16.853	111.738	130.425,95	77,36	77,39
Kab. Lima Puluh Kota	4.776	6.129	30.252	38.839,8	63,34	63,37
Kab. Pasaman	12.677	11.615,8	79.690	73.063,38	62,86	62,9
Kab. Solok Selatan	14.452	15.308,2	89.539	94.878,27	61,96	61,98
Kab. Dharmasraya	3.824	2.498,8	23.357	15.270,66	61,08	61,11
Kab. Pasaman Barat	53.921	56.863,8	340.916	364.291	63,23	64,06
Kota Padang	0	74,3	0	0	0	0
Kota Solok	45	66,5	236	351,23	52,44	52,82
Kota Sawahlunto	5	35,5	36	241,23	72	67,95
Kota Padang Panjang	0	0	0	0	0	0
Kota Bukittinggi	17	9,9	63	37,36	37,06	37,74
Kota Payakumbuh	246	306,5	1.414	1.759,58	57,48	57,41
Kota Pariaman	40	40	204	204,08	51	51,02

Sumber : Badan Pusat Statistik (BPS)

**Tabel 2. Data Tinggi Wilayah Sumatera Barat**

Kabupaten/Kota	Tinggi Wilayah (mdpl)
Kab. Kepulauan Mentawai	2
Kab. Pesisir Selatan	5
Kab. Solok	1.006
Kab. Sijunjung	160
Kab. Tanah Datar	460
Kab. Padang Pariaman	2
Kab. Agam	102
Kab. Lima Puluh Kota	513
Kab. Pasaman	450
Kab. Solok Selatan	350
Kab. Dharmasraya	115
Kab. Pasaman Barat	3
Kota Padang	2
Kota Solok	390
Kota Sawahlunto	262



Kota Padang Panjang	780
Kota Bukittinggi	927
Kota Payakumbuh	515
Kota Pariaman	2

Sumber : Badan Pusat Statistik (BPS)

### 3.2.1.2 Data Variabel Curah Hujan

Data primer dalam penelitian ini diperoleh dari Stasiun Klimatologi Badan Meteorologi, Klimatologi, dan Geofisika (Staklim BMKG) Sicincin. Data curah hujan disajikan dalam Tabel 3.

Tabel 3. Data Curah Hujan Sumatera Barat

Kabupaten/Kota	Curah Hujan (mm/tahun)	
	2017	2018
Kab. Kepulauan Mentawai	3.446	2.169
Kab. Pesisir Selatan	4.382	4.297
Kabupaten/Kota	Curah Hujan (mm/tahun)	
	2017	2018
Kab. Solok	1.934	2.304
Kab. Sijunjung	2.547	2.333
Kab. Tanah Datar	1.774	1.935
Kab. Padang Pariaman	4.270	4.165
Kab. Agam	3.947	3.047
Kab. Lima Puluh Kota	2.579	2.408
Kab. Pasaman	3.737	4.077
Kab. Solok Selatan	4.361	2.600
Kab. Dharmasraya	2.640	2.909
Kab. Pasaman Barat	4.915	4.689
Kota Padang	4.879	4.416
Kota Solok	1.821	1.827
Kota Sawahlunto	2.014	2.115
Kota Padang Panjang	3.501	2.634
Kota Bukittinggi	2.104	2.696
Kota Payakumbuh	2.026	2.606
Kota Pariaman	3.224	4.313

Sumber : Staklim BMKG

### 3.2.2 Preprocessing Data

Preprocessing data merupakan tahapan awal yang sangat krusial dalam proses ekstraksi pola atau informasi penting, sebab data yang tidak bersih dan tidak terstandarisasi dapat menurunkan kualitas hasil analisis [16], [17]. Dalam penelitian sentra produksi jagung (*Zea mays L.*) di Sumatera Barat, data sekunder yang diperoleh dari instansi terkait sering kali tidak lengkap, seperti terdapat data hilang (*missing values*) atau tidak valid. Selain itu, beberapa atribut data yang tidak relevan perlu dibuang melalui proses seleksi karena dapat mengurangi akurasi hasil pengelompokan (*Clustering*). Data deret waktu tahunan (2017-2022) tersebut kemudian diekstraksi baik dalam bentuk tabel maupun hasil olahan perangkat lunak



menjadi basis data terstruktur untuk mempermudah pengolahan. Tahapan pra-pemrosesan data dalam penelitian ini meliputi:

### 3.2.2.1 Data Selection

Seleksi data merupakan langkah awal dalam menetapkan variabel yang akan digunakan sebagai *dataset* dalam proses pengelompokan menggunakan algoritma *K-means*. Berdasarkan sejumlah sumber data yang tersedia, dilakukan pemilihan variabel utama (*input*) dan variabel tambahan (pendukung) yang relevan dengan kajian. Variabel *input* yang digunakan antara lain meliputi luas panen jagung, produksi jagung, produktivitas jagung, tinggi wilayah, dan curah hujan. Adapun variabel pendukung yang disertakan meliputi nama kabupaten atau kota, dan tahun pengamatan. Data hasil seleksi tersebut disajikan dalam bentuk Tabel 4.

Tabel 4. Data Selection

Atribut	Keterangan	Jenis Data
<b>Indikator Input</b>		
Luas Panen	Total luas lahan yang ditanami jagung dan siap dipanen dalam satuan hektare (ha).	Numerik
Produksi Jagung	Jumlah total hasil panen jagung dalam satuan ton.	Numerik
Produktivitas Jagung	Hasil panen jagung per satuan luas lahan (biasanya dalam kuintal/hektare).	Numerik
<b>Atribut</b>		
Tinggi Wilayah	Ketinggian lokasi penanaman jagung di atas permukaan laut (mdpl).	Numerik
Curah Hujan	Jumlah presipitasi (hujan) tahunan dalam satuan milimeter (mm/tahun).	Numerik
<b>Indikator Pendukung</b>		
Nama Kabupaten/Kota	Label wilayah administrasi tempat data diambil.	Nominal
Tahun	Periode waktu pengamatan data (2017- 2022).	Nominal

Sumber : Badan Pusat Statistik (BPS)

### 3.2.2.2 Data Integration

Tahap integrasi data pada penelitian ini dilakukan dengan menggabungkan data yang berasal dari berbagai sumber query menjadi satu kesatuan dalam bentuk basis data yang seragam dan siap diolah. Proses integrasi data harus dilakukan dengan ketelitian tinggi, karena kesalahan dalam penggabungan data dapat menyebabkan hasil analisis yang menyesatkan. Sebagai ilustrasi, apabila penggabungan data dilakukan berdasarkan kategori yang tidak sejenis, seperti mencampurkan data dari jenis komoditas berbeda, maka dapat muncul korelasi yang keliru dan mempengaruhi akurasi pengambilan keputusan [9]. Proses integrasi dari berbagai sumber data dapat dilihat secara rinci pada bagian lampiran. Data hasil penggabungan tersebut merupakan informasi dari berbagai instansi di Sumatera Barat pada tahun 2017. Adapun variabel yang digunakan meliputi: X1 (luas panen), X2 (produksi jagung), X3 (produktivitas jagung), X4 (tinggi wilayah), dan X5 (curah hujan). Data tersebut disajikan dalam Tabel 5.



**Tabel 5. Database Penelitian Tahun 2017**

Kabupaten/Kota	x1 luas panen (hektar)	x2 produksi (ton)	x3 produkivitas (Kuintal/t)	x4 tinggi wilayah (mdpl)	x5 curah hujan
Kab. Kepulauan Mentawai	2	10	50	2	3.446
Kab. Pesisir Selatan	20.829	169.102	81,19	5	4.382
Kab. Solok	397	4.018	101,21	1.006	1.934
Kab. Sijunjung	456	2.956	64,82	160	2.547
Kab. Tanah Datar	6.804	52.046	76,49	460	1.774
Kab. Padang Pariaman	9.400	80.270	85,39	2	4.270
Kab. Agam	14.444	111.738	77,36	102	3.947
Kab. Lima Puluh Kota	4.776	30.252	63,34	513	2.579
Kab. Pasaman	12.677	79.690	62,86	450	3.737
Kab. Solok Selatan	14.452	89.539	61,96	350	4.361
Kab. Dharmasraya	3.824	23.357	61,08	115	2.640
Kab. Pasaman Barat	53.921	340.916	63,23	3	4.915
Kota Padang	0	0	0	2	4.879
Kota Solok	45	236	52,44	390	1.821
Kota Sawahlunto	5	36	72	262	2.014
Kota Padang Panjang	0	0	0	780	3.501
Kota Bukittinggi	17	63	37,06	927	2.104
Kota Payakumbuh	246	1.414	57,48	515	2.026
Kota Pariaman	40	204	51	2	3.224

### 3.2.2.3 Data Transformasi

Transformasi data merupakan proses untuk mengubah skala nilai pada masing-masing variabel agar sesuai dengan kebutuhan analisis menggunakan algoritma *K-means*. Perbedaan rentang nilai yang terlalu jauh antar variabel dapat menyulitkan proses pengelompokan, karena berpotensi menimbulkan bias dan ketidak seimbangan dalam perhitungan jarak antar data. Oleh karena itu, apabila terdapat perbedaan satuan atau skala yang signifikan, maka dilakukan standarisasi data menggunakan rumus Z-score guna menciptakan kesetaraan antar variabel. Proses transformasi ini dijelaskan pada Persamaan 4, sedangkan hasil standarisasi data disajikan pada Tabel 6.

**Tabel 6. Menghitung Standar Deviasi Luas Panen (X1) Tahun 2017**

No.	Kabupaten/Kota	X1 2017	(X1- $\bar{X}$ ) 2017	(X1- $\bar{X}$ ) <sup>2</sup> 2017
1.	Kab. Kepulauan Mentawai	2	-7.489,32	56.089.850,99
2.	Kab. Pesisir Selatan	20.829	13.337,68	177.893.820,1
3.	Kab. Solok	397	-7.094,32	50.329.316,52
4.	Kab. Sijunjung	456	-7.035,32	49.495.668,26
5.	Kab. Tanah Datar	6.804	-687,32	472.402,99
6.	Kab. Padang Pariaman	9.400	1.908,68	3.643.075,42
7.	Kab. Agam	14.444	6.952,68	48.339.817,73
8.	Kab. Lima Puluh Kota	4.776	-2.715,32	7.372.939,84
9.	Kab. Pasaman	12.677	5.185,68	26.891.320,73



10.	Kab. Solok Selatan	14.452	6.960,68	48.451.124,68
11.	Kab. Dharmasraya	3.824	-3.667,32	13.449.205,1
12.	Kab. Pasaman Barat	53.921	46.429,68	2.155.715.576
13.	Kota Padang	0	-7.491,32	56.119.812,26
14.	Kota Solok	45	-7.446,32	55.447.618,84
15.	Kota Sawahlunto	5	-7.486,32	56.044.924,1
16.	Kota Padang Panjang	0	-7.491,32	56.119.812,26
17.	Kota Bukittinggi	17	-7.474,32	55.865.396,52
18.	Kota Payakumbuh	246	-7.245,32	52.494.600,89
19.	Kota Pariaman	40	-7.451,32	55.522.106,99
Total		142.335		3.025.758.390
rata-rata		7.491,316		302.575.839
V				168.097.688,3
SD				12.965,24926

Hasil transformasi data untuk masing-masing variabel sentra produksi jagung di Provinsi Sumatera Barat tahun 2017 disajikan pada Tabel 7.

Tabel 7. Database Z-score Tahun 2017

No.	Kabupaten/Kota	Luas Panen	Produksi	Produktivitas	Tinggi Wilayah	Curah Hujan	
			X1	X2	X3	X4	X5
Kab.Kepulauan Mentawai							
1		-0,57765	-0,60773	-0,35149	-0,97562	0,26013	
2	Kab. Pesisir Selatan	1,02873	1,37316	0,88169	-0,96636	1,12114	
3	Kab. Solok	-0,54718	-0,56077	1,67324	2,12207	-1,13072	
4	Kab. Sijunjung	-0,54263	-0,57321	0,23446	-0,48813	-0,56684	
5	Kab. Tanah Datar	-0,05301	0,00187	0,69587	0,43747	-1,27790	
6	Kab. Padang Pariaman	0,14722	0,33251	1,04775	-0,97562	1,01811	
7	Kab. Agam	0,53626	0,70115	0,73026	-0,66708	0,72099	
8	Kab. Lima Puluh Kota	-0,20943	-0,25345	0,17594	0,60099	-0,53740	
9	Kab. Pasaman	0,39997	0,32571	0,15697	0,40662	0,52782	
No.	Kabupaten/Kota	Luas Panen	Produksi	Produktivitas	Tinggi Wilayah	Curah Hujan	
			X1	X2	X3	X4	X5
10	Kab. Solok Selatan		0,53687	0,44109	0,12138	0,09808	1,10182
11	Kab. Dharmasraya		-0,28286	-0,33422	0,08659	-0,62698	-0,48129
12	Kab. Pasaman Barat		3,58109	3,38593	0,17159	-0,97253	1,61143
13	Kota Padang		-0,57780	-0,60784	-2,32838	-0,97562	1,57832
14	Kota Solok		-0,57433	-0,60508	-0,25502	0,22150	-1,23467
15	Kota Sawahlunto		-0,57741	-0,60742	0,51834	-0,17343	-1,05713
16	Kota Padang Panjang		-0,57780	-0,60784	-2,32838	1,42478	0,31073
17	Kota Bukittinggi		-0,57649	-0,60711	-0,86311	1,87833	-0,97434



18	Kota Payakumbuh	-0,55883	-0,59128	-0,05575	0,60716	-1,04610
19	Kota Pariaman	-0,57471	-0,60545	-0,31195	-0,97562	0,05592

### 3.3 Proses Clustering Menggunakan Algoritma K-means

Setelah data melalui tahapan *preprocessing*, langkah selanjutnya adalah menerapkan metode *K-means* untuk proses pengelompokan. Alur tahapan dalam algoritma ini dijelaskan pada *flowchart*. Sebagai contoh penerapan, peneliti menggunakan data dari salah satu tahun dari 2017 hingga 2022 untuk menggambarkan proses kerja algoritma *K-means*, dengan ilustrasi data ditampilkan pada tabel *database z-score* tahun 2017. Adapun langkah-langkah pelaksanaan *K-means* dijelaskan sebagai berikut :

#### 3.3.1 Menentukan Jumlah Cluster

Dalam penerapan algoritma *K-means*, tahap awal yang harus dilakukan adalah menetapkan jumlah *Cluster* yang akan digunakan. Pada penelitian ini, jumlah *Cluster* ditentukan secara bertahap, mulai dari 2 hingga 6 *Cluster*. Setiap konfigurasi *Cluster* tersebut digunakan sebagai dasar dalam proses pengambilan keputusan. Untuk mengetahui jumlah *Cluster* yang paling optimal, dilakukan pengujian menggunakan metode *Elbow*, yang mampu menunjukkan titik terbaik berdasarkan nilai *Sum of Squared Error* (SSE) dari masing-masing skenario jumlah *Cluster*.

#### 3.3.2 Menentukan Titik Pusat Cluster atau Centroid

Penentuan pusat *Cluster* didasarkan pada variabel-variabel sentra produksi jagung yang telah disesuaikan dengan jumlah *Cluster* yang digunakan. Titik pusat (*centroid*) ditentukan secara acak sebagai langkah awal dalam proses iterasi. Sebagai contoh, pada data tahun 2017, variabel yang digunakan meliputi Y1 (luas panen), Y2 (produksi jagung), Y3 (produktivitas jagung), Y4 (tinggi wilayah), dan Y5 (curah hujan). Nilai-nilai awal pusat *cluster* yang diperoleh ditampilkan pada Tabel 8.

Tabel 8. Database Pusat Cluster Tahun 2017

Jumlah Cluster	Kelompok Cluster	Data Z-score				
		Luas Panen Y1	Produksi Y2	Produktivitas Y3	Tinggi Wilayah Y4	Curah Hujan Y5
2 Cluster	1	-0,54263	-0,57321	0,23446	-0,48813	-0,56684
	2	-0,57741	-0,60742	0,51834	-0,17343	-1,05713
3 Cluster	1	-0,54718	-0,56077	1,67324	2,12207	-1,13072
	2	0,14722	0,33251	1,04775	-0,97562	1,01811
	3	-0,57741	-0,60742	0,51834	-0,17343	-1,05713
4 Cluster	1	-0,54718	-0,56077	1,67324	2,12207	-1,13072
	2	0,14722	0,33251	1,04775	-0,97562	1,01811
	3	-0,57780	-0,60784	-2,32838	-0,97562	1,57832
	4	-0,55883	-0,59128	-0,05575	0,60716	-1,04610
5 Cluster	1	-0,57765	-0,60773	-0,35149	-0,97562	0,26013
	2	-0,54263	-0,57321	0,23446	-0,48813	-0,56684
	3	0,53687	0,44109	0,12138	0,09808	1,10182
Jumlah Cluster	Kelompok Cluster	Data Z-score				
		Luas Panen Y1	Produksi Y2	Produktivitas Y3	Tinggi Wilayah Y4	Curah Hujan Y5
	4	-0,57433	-0,60508	-0,25502	0,22150	-1,23467



6 Cluster	5	-0,55883	-0,59128	-0,05575	0,60716	-1,04610
	1	-0,57765	-0,60773	-0,35149	-0,97562	0,26013
	2	-0,54718	-0,56077	1,67324	2,12207	-1,13072
	3	0,14722	0,33251	1,04775	-0,97562	1,01811
	4	-0,28286	-0,33422	0,08659	-0,62698	-0,48129
	5	-0,57433	-0,60508	-0,25502	0,22150	-1,23467
	6	-0,57471	-0,60545	-0,31195	-0,97562	0,05592

### 3.3.3 Menghitung Jarak ke Titik Pusat Cluster

Proses pengelompokan data ke dalam masing-masing *Cluster* dilakukan dengan terlebih dahulu menghitung jarak antara setiap data dan pusat *Cluster*. Penghitungan jarak ini menggunakan rumus *Euclidean distance*, sebagaimana dijelaskan pada bagian sebelumnya. Sebagai ilustrasi, dilakukan perhitungan jarak untuk mengukur kemiripan Kabupaten Kepulauan Mentawai terhadap tiga pusat *cluster* pada data tahun 2017.

#### 3.3.3.1 Iterasi Pertama

Proses iterasi dilakukan untuk mengamati jarak kedekatan masing-masing data terhadap pusat *cluster*. Hasil perhitungan jarak pada iterasi pertama disajikan pada Tabel 9.

Tabel 9. Jarak Data Mentawai ke *Centroid*

Objek	X1	X2	X3	X4	X5
Mentawai (X)	-0,57765	-0,60773	-0,35149	-0,97562	0,26013
Centroid1 (Y)	-0,54718	-0,56077	1,67324	2,12207	-1,13072
Centroid2 (Y)	0,14722	0,33251	1,04775	-0,97562	1,01811
Centroid3 (Y)	-0,57741	-0,60742	0,51834	-0,17343	-1,05713
C1 (X-Y)	-0,03047	-0,04695	-2,02473	-3,09769	1,39086
C2 (X-Y)	-0,72486	-0,94023	-1,39924	0,00000	-0,75798
C3 (X-Y)	-0,00023	-0,00030	-0,86983	-0,80219	1,31727
C1 (X-Y) <sup>2</sup>	0,00093	0,00220	4,09953	9,59568	1,93448
C2 (X-Y) <sup>2</sup>	0,52542	0,88404	1,95788	0,00000	0,57453
C3 (X-Y) <sup>2</sup>	0,00000	0,00000	0,75661	0,64351	1,73519

Tahapan perhitungan jarak antar data dapat dijelaskan sebagai berikut.

#### 1. Jarak Cluster 1

$$D_{ij} = \sqrt{\sum_{i=1}^p (X - Y)^2}$$

$$D_{ij} = \sqrt{(-0,57765) - (-0,54718))^2 + (-0,60773) - (-0,56077))^2 + ((-0,35149) - (1,67324))^2 + ((-0,97562) - (2,12207))^2 + ((0,26013) - (1,01811))^2}$$

$$D_{ij} = 3,95384$$

#### 2. Jarak Cluster 2

$$D_{ij} = \sqrt{\sum_{i=1}^p (X - Y)^2}$$

$$D_{ij} = \sqrt{(-0,57765) - (0,14722))^2 + (-0,60773) - (0,33251))^2 + ((-0,35149) - (1,04775))^2 + ((-0,97562) - (-0,97562))^2 + ((0,26013) - (-1,13072))^2}$$

$$D_{ij} = 1,98542$$



### 3. Jarak Cluster

$$D_{ij} = \sqrt{\sum_{i=1}^p (X_i - Y_i)^2}$$

$$D_{ij} = \sqrt{(-0,57765) - (-0,57741))^2 + (-0,60773) - (-0,60742))^2 + ((-0,35149) - (0,51834))^2 + ((-0,97562) - (-0,17343))^2 + ((0,26013) - (-1,05713))^2}$$

$$D_{ij} = 1,77068$$

Berdasarkan data Kabupaten Kepulauan Mentawai pada baris pertama, diperoleh hasil perhitungan jarak terhadap pusat Cluster yaitu C1 sebesar 3,95384, C2 sebesar 1,98542, dan C3 sebesar 1,77068. Langkah selanjutnya adalah menentukan nilai jarak terkecil untuk mengidentifikasi Cluster tempat Kabupaten Kepulauan Mentawai dikelompokkan. Untuk menentukan keanggotaan Cluster, digunakan pendekatan nilai minimum dari ketiga jarak tersebut. Karena jarak terkecil adalah 1,77068 maka Kabupaten Kepulauan Mentawai dikelompokkan ke dalam Cluster 3. Perhitungan jarak antar data dengan pusat cluster pada iterasi pertama dapat dilihat pada Tabel 10.

Tabel 10. Hasil Perhitungan Jarak Iterasi Pertama

No.	kabupaten	D1	D2	D3	Cluster
1	Kab. Kepulauan Mentawai	3,95384	1,98541	1,77068	3
2	Kab. Pesisir Selatan	4,63243	1,37779	3,46526	2
3	Kab. Solok	0,00000	3,98554	2,57131	1
4	Kab. Sijunjung	3,03338	2,16965	0,64992	3
5	Kab. Tanah Datar	2,09178	2,74622	1,04866	3
6	Kab. Padang Pariaman	3,98554	0,00000	2,57662	2
7	Kab. Agam	3,85535	0,75599	2,53040	2
8	Kab. Lima Puluh Kota	2,26189	2,47708	1,11705	3
9	Kab. Pasaman	3,11056	1,73447	2,19206	2
10	Kab. Solok Selatan	3,69702	1,47703	2,68950	2
11	Kab. Dharmasraya	3,25848	1,98068	0,94080	3
12	Kab. Pasaman Barat	7,20902	4,71535	6,41249	2
13	Kota Padang	5,74027	3,62243	3,96143	2
14	Kota Solok	2,70996	3,09920	0,88633	3
15	Kota Sawahlunto	2,57131	2,57662	0,00000	3
16	Kota Padang Panjang	4,31046	4,36698	3,53965	3
17	Kota Bukittinggi	2,55342	4,14402	2,47487	3
18	Kota Payakumbuh	2,30056	3,05544	0,96935	3
19	Kota Pariaman	3,86621	2,04342	1,60368	3

Berdasarkan hasil perhitungan jarak pada iterasi pertama, diketahui bahwa jumlah anggota pada masing-masing *cluster* adalah sebagai berikut: *Cluster 1* terdiri dari 1 anggota, *Cluster 2* terdiri dari 2 anggota, dan *Cluster 3* terdiri dari 11 anggota. Setelah diketahui komposisi anggota pada masing-masing *cluster* pada iterasi pertama, langkah selanjutnya adalah menghitung ulang titik pusat (*centroid*) berdasarkan data yang tergabung dalam *cluster* yang sama. Perhitungan pusat *cluster* terbaru disajikan pada bagian berikut:

$$\text{Centroid} = n \sum m$$

$$C1X1 = 1 - 0,54718 = -0,54718$$



Perhitungan dilakukan untuk masing-masing variabel X1 hingga X5 pada setiap *centroid*, yaitu *centroid 1*, *centroid 2*, dan *centroid 3*. Nilai-nilai *centroid* yang dihasilkan dari proses ini akan digunakan sebagai pusat *cluster* pada iterasi kedua. Proses perhitungan dalam algoritma *K-means* dilakukan secara berulang melalui tahapan iterasi. Setiap iterasi terdiri dari perhitungan jarak antara data dan pusat *cluster*, penentuan keanggotaan *cluster* berdasarkan jarak terdekat, dan perhitungan ulang pusat *cluster* berdasarkan data yang tergabung dalam masing-masing kelompok. Langkah-langkah tersebut terus diulang sebagaimana telah dijelaskan pada bagian sebelumnya. Proses iterasi ini akan berhenti ketika tidak terjadi lagi perubahan anggota pada masing-masing cluster, atau dengan kata lain, ketika hasil pengelompokan telah mencapai kondisi konvergen.

### 3.3.4 Menentukan Jumlah *Cluster* Terbaik

Salah satu pendekatan yang digunakan untuk menentukan jumlah *cluster* optimal adalah metode Elbow. Metode ini mengidentifikasi titik terbaik berdasarkan pola penurunan nilai SSE (*Sum of Squared Error*), yang akan membentuk sudut seperti siku pada grafik. Titik tersebut menjadi indikasi jumlah *cluster* yang paling sesuai untuk digunakan dalam proses *Clustering* [18], [19]. Dalam penelitian ini, jumlah *cluster* yang diuji dimulai dari 2 hingga 6 *cluster*. Untuk masing-masing jumlah *cluster* tersebut, dilakukan perhitungan nilai (SSE) berdasarkan hasil pengelompokan pada iterasi terakhir. Nilai SSE dihitung menggunakan rumus yang mengukur total jarak kuadrat antara data dan *centroid* clusternya. Langkah-langkah perhitungan SSE dijelaskan pada bagian berikut.

Langkah pertama dalam menghitung nilai SSE adalah dengan mengamati hasil pengelompokan pada iterasi terakhir. Proses ini dilakukan dengan menandai nilai jarak terkecil pada setiap baris data untuk menentukan keanggotaan clusternya, kemudian mengabaikan nilai jarak terhadap *cluster* lain yang tidak relevan. Rincian hasil tersebut disajikan pada Tabel 11.

Tabel 11. Pengelompokan *Cluster* Jarak Minimum

No.	Kabupaten/Kota	D1	D2	D3	Cluster
1	Kab. Kepulauan Mentawai			1,449182	3
2	Kab. Pesisir Selatan		1,032522		2
3	Kab. Solok	0			1
4	Kab. Sijunjung			0,814671	3
5	Kab. Tanah Datar			1,337939	3
6	Kab. Padang Pariaman	1,320589			2
7	Kab. Agam	0,792302			2
8	Kab. Lima Puluh Kota		0,681563		3
9	Kab. Pasaman	1,318571			2
10	Kab. Solok Selatan	0,83612			2
11	Kab. Dharmasraya		0,900024		3
12	Kab. Pasaman Barat	3,813882			2
13	Kota Padang	3,222964			2
14	Kota Solok		0,661234		3
15	Kota Sawahlunto		0,954664		3
16	Kota Padang Panjang		2,615443		3
17	Kota Bukittinggi		1,86525		3
18	Kota Payakumbuh		0,66086		3
19	Kota Pariaman		1,335162		3

Tahap selanjutnya adalah menghitung nilai SSE menggunakan rumus yang telah dijelaskan sebelumnya. Hasil perhitungan nilai SSE tersebut disajikan pada Tabel 12.



Tabel 12. Perhitungan SSE untuk Tiga Cluster pada Tahun 2017

No.	Kabupaten/Kota	D1	D2	D3	Cluster
1	Kab. Kepulauan Mentawai			2,10013	3
2	Kab. Pesisir Selatan		1,066102		2
3	Kab. Solok	0			1
4	Kab. Sijunjung			0,663689	3
5	Kab. Tanah Datar			1,790081	3
6	Kab. Padang Pariaman	1,743954			2
7	Kab. Agam	0,627743			2
8	Kab. Lima Puluh Kota			0,464528	3
No.	Kabupaten/Kota	D1	D2	D3	Cluster
9	Kab. Pasaman		1,73863		2
10	Kab. Solok Selatan	0,699096			2
11	Kab. Dharmasraya			0,810044	3
12	Kab. Pasaman Barat	14,5457			2
13	Kota Padang	10,38749			2
14	Kota Solok		0,43723		3
15	Kota Sawahlunto		0,911384		3
16	Kota Padang Panjang		6,84054		3
17	Kota Bukittinggi		3,479158		3
18	Kota Payakumbuh		0,436736		3
19	Kota Pariaman		1,782657		3
Jumlah SSE		0	30,80872	19,71618	
Total SSE		50,52489			

Tabel 12 di atas menampilkan hasil perhitungan nilai SSE untuk tahun 2017 dengan jumlah cluster sebanyak tiga. Prosedur yang sama juga diterapkan untuk masing-masing tahun pada setiap variasi jumlah cluster, mulai dari dua hingga enam cluster. Rekapitulasi total nilai SSE secara keseluruhan dapat dilihat pada Tabel 13.

Tabel 13. Total Nilai SSE dari Tahun 2017 – 2022

Tahun	Total Nilai SSE Tiap Cluster				
	2	3	4	5	6
2017	59,8759	50,52489	34,56591	34,31633	24,50082
2018	59,48546	45,39558	28,24574	29,16887	23,54907
2019	59,38179	46,16891	38,1809	24,11762	23,0253
2020	54,48292	42,89165	28,78836	18,02762	14,2911
2021	66,54783	62,48695	40,861	30,24664	21,49652
2022	67,62959	39,14898	32,90212	32,37931	19,39616

Berdasarkan Tabel 13 di atas, terlihat bahwa peningkatan jumlah Cluster cenderung menyebabkan penurunan nilai SSE. Penentuan jumlah cluster yang paling optimal dilakukan dengan mengidentifikasi selisih terbesar antar nilai SSE yang dihasilkan. Rincian nilai selisih SSE tersebut disajikan pada Tabel 14.



Tabel 14. Selisih Nilai SSE dari Tahun 2017 – 2022

Tahun	Selisih Nilai SSE			
	2 ke 3	3 ke 4	4 ke 5	5 ke 6
2017	9,35100	15,95898	0,24958	9,81551
2018	14,08988	17,14984	0,92313	5,61980
2019	13,21288	7,98801	14,06328	1,09232
2020	11,59127	14,10329	10,76073	3,73652
2021	4,06088	21,62595	10,61435	8,75013
2022	28,48061	6,24685	0,52282	12,98315

Berdasarkan Tabel 14 di atas, dapat disimpulkan bahwa jumlah *Cluster* optimal pada masing-masing tahun, yang ditentukan berdasarkan nilai selisih SSE terbesar selama periode 2017 hingga 2022, secara berturut-turut adalah sebagai berikut, 15,95898; 17,14984; 14,0633; 14,10329; 21,62595; 28,48061. Penentuan jumlah *cluster* terbaik didasarkan pada nilai SSE yang menunjukkan penurunan paling signifikan dan membentuk pola grafik menyerupai siku, sebagaimana yang diilustrasikan dalam metode Elbow. Sebagai contoh, pada periode tahun 2017, selisih nilai SSE terbesar terjadi antara jumlah *cluster* 3 ke *cluster* 4, yaitu sebesar 15,95898. Oleh karena itu, dapat disimpulkan bahwa jumlah *cluster* terbaik pada tahun 2017 adalah 4 *cluster*. Berdasarkan hasil yang diperoleh selama periode 2017 hingga 2022, jumlah *cluster* optimal secara berurutan adalah: 4 *cluster*, 4 *cluster*, 5 *cluster*, 4 *cluster*, 4 *cluster*, dan 3 *cluster*.

### 3.3.5 Identifikasi Wilayah Sentra Produksi Jagung

Penentuan daerah yang direkomendasikan sebagai sentra produksi jagung dilakukan berdasarkan hasil akhir pengelompokan data menggunakan algoritma *K-means Clustering*. Setelah melalui tahapan *preprocessing*, standarisasi, proses iterasi, serta penentuan jumlah *cluster* terbaik dengan metode *Elbow*, hasil *Clustering* yang paling representatif digunakan sebagai dasar dalam menyusun rekomendasi wilayah sentra produksi. Selanjutnya, penentuan wilayah yang direkomendasikan sebagai sentra produksi jagung dari tahun 2017 hingga 2022 didasarkan pada nilai pusat *Cluster* akhir dari hasil *Cluster* terbaik. Rincian nilai pusat *cluster* tersebut ditampilkan pada Tabel 15.

Tabel 15. Rekapitulasi Titik Pusat Cluster Terbaik

Tahun	Cluster	Luas Panen	Produksi	Produktivitas	Tinggi Wilayah	Curah Hujan
		Y1	Y2	Y3	Y4	Y5
2017	1	-0,54718	-0,56077	1,67324	2,12207	-1,13072
	2	1,03835	1,09326	0,51827	-0,51282	1,01688
	3	-0,57780	-0,60784	-2,32838	0,22458	0,94452
	4	-0,45273	-0,47831	-0,01261	0,05057	-0,68596
2018	1	0,26449	0,37269	0,50880	-0,63561	0,99124
	2	-0,33023	-0,36037	0,27544	0,18231	-0,73747
	3	3,64872	3,49767	0,20952	-0,97253	1,73015
	4	-0,55630	-0,58581	-1,83597	0,77583	0,22947
2019	1	-0,55328	-0,59254	-0,33556	-0,59201	-0,70135
	2	-0,40852	-0,42875	0,46177	0,97792	-0,80113
	3	3,54661	3,35104	-0,16294	-0,97253	1,40950
	4	-0,57771	-0,61636	-3,17850	1,42478	-0,62383
2020	5	0,14276	0,20193	0,19719	-0,56797	0,76564
	1	-0,51705	-0,52734	-0,10474	-0,62119	0,04096
	2	1,58693	1,63526	0,34963	-0,62698	1,18668
	3	-0,26671	-0,28305	0,43631	1,00877	-0,83030



	4	-0,61102	-0,62395	-3,17850	1,42478	-0,09258
2021	1	-0,15569	-0,08590	0,31061	-0,91597	1,15254
	2	-0,47618	-0,47724	-0,52520	1,17847	-0,53477
	3	1,38636	1,27681	0,20359	-0,28373	-1,00560
	4	-0,58474	-0,57613	0,15772	-0,14669	0,10527
2022	1	-0,51301	-0,51855	0,05131	-0,90589	1,19582
	2	-0,53737	-0,52627	-0,28324	0,73675	-0,46608
	3	1,48028	1,46583	0,45852	-0,42026	-0,35688

Berdasarkan nilai akhir pusat *cluster* (centroid) dari masing-masing variabel hasil *K-Means Clustering*, diketahui bahwa terdapat nilai *Z-score* yang bersifat negatif maupun positif. Nilai *Z-score* positif mengindikasikan bahwa suatu variabel berada di atas rata-rata keseluruhan, sedangkan nilai negatif menunjukkan berada di bawah rata-rata. Klaster yang memiliki nilai rata-rata tertinggi (positif) pada masing-masing variabel, seperti luas panen, produksi, produktivitas, tinggi wilayah, dan curah hujan diidentifikasi sebagai klaster unggulan, karena menggambarkan wilayah dengan karakteristik produksi jagung yang relatif lebih baik dibanding wilayah lain. Selanjutnya, dilakukan perbandingan hasil klasterisasi dari tahun 2017 hingga 2022 untuk melihat konsistensi wilayah yang tergabung dalam klaster unggulan. Berdasarkan Tabel 15, dapat diketahui bahwa jumlah klaster optimal yang diperoleh dari metode Elbow untuk setiap tahun analisis adalah sebagai berikut: tahun 2017 sebanyak 4 klaster, 2018 sebanyak 4 klaster, 2019 sebanyak 5 klaster, 2020 sebanyak 4 klaster, 2021 sebanyak 4 klaster, dan 2022 sebanyak 3 klaster. Selanjutnya, pada masing-masing tahun dilakukan identifikasi terhadap centroid yang memiliki nilai positif terbanyak pada lima variabel penelitian, yaitu luas panen, produksi, produktivitas, ketinggian wilayah, dan curah hujan. Centroid dengan dominasi nilai positif pada variabel-variabel tersebut dianggap merepresentasikan kelompok wilayah yang memiliki karakteristik unggul dan berpotensi sebagai sentra produksi jagung.

Hasil identifikasi menunjukkan bahwa:

- Pada tahun 2017, centroid terbaik adalah centroid 2 dengan anggota: Kabupaten Pesisir Selatan, Kabupaten Padang Pariaman, Kabupaten Agam, Kabupaten Pasaman, dan Kabupaten Solok Selatan.
- Tahun 2018, terdapat dua centroid unggulan, yaitu centroid 1 dan centroid 3, yang meliputi: Kabupaten Pesisir Selatan, Kabupaten Padang Pariaman, Kabupaten Agam, Kabupaten Pasaman, Kota Pariaman, dan Kabupaten Pasaman Barat.
- Tahun 2019, centroid 5 menjadi klaster unggulan dengan anggota: Kabupaten Pesisir Selatan, Kabupaten Sijunjung, Kabupaten Padang Pariaman, Kabupaten Agam, Kabupaten Pasaman, Kabupaten Solok Selatan, Kota Padang, dan Kota Pariaman.
- Tahun 2020, centroid 2 memiliki nilai positif terbanyak, dengan anggota: Kabupaten Pesisir Selatan, Kabupaten Agam, Kabupaten Solok Selatan, dan Kabupaten Pasaman Barat.
- Tahun 2021, centroid 3 terdiri atas: Kabupaten Agam, Kabupaten Pasaman, Kabupaten Solok Selatan, dan Kabupaten Pasaman Barat.
- Tahun 2022, centroid 3 kembali menjadi klaster unggulan dengan anggota: Kabupaten Pesisir Selatan, Kabupaten Agam, Kabupaten Pasaman, Kabupaten Solok Selatan, dan Kabupaten Pasaman Barat.

Berdasarkan hasil identifikasi tersebut, dilakukan rekapitulasi frekuensi kemunculan setiap daerah dalam klaster unggulan pada rentang tahun 2017 hingga 2022. Daerah-daerah yang paling konsisten muncul dalam klaster dengan nilai centroid tertinggi dianggap sebagai daerah yang memiliki potensi kuat sebagai sentra produksi jagung [11]. Dari hasil rekapitulasi tersebut, diperoleh lima daerah yang paling sering muncul, yaitu: Kabupaten Pesisir Selatan, Kabupaten Agam, Kabupaten Pasaman, Kabupaten Solok Selatan, dan Kabupaten Pasaman Barat. Kelima daerah tersebut direkomendasikan sebagai daerah sentra produksi jagung di Provinsi Sumatera Barat berdasarkan hasil analisis data multivariat dan pemodelan klaster *K-Means* selama enam tahun terakhir.



## 4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai penentuan daerah sentra produksi jagung (*Zea mays L.*) di Kabupaten/Kota Provinsi Sumatera Barat menggunakan metode *K-Means Clustering*, dapat disimpulkan beberapa hal sebagai berikut:

- 1 Penelitian ini telah mengidentifikasi dan menentukan wilayah-wilayah potensial sebagai sentra produksi jagung di Provinsi Sumatera Barat dengan menggunakan pendekatan analisis multivariat. Pengelompokan dilakukan terhadap 19 kabupaten/kota berdasarkan lima variabel utama, yaitu: luas panen, produksi, produktivitas, ketinggian wilayah, dan curah hujan.
- 2 Hasil analisis menunjukkan adanya pengelompokan wilayah ke dalam beberapa klaster berdasarkan kemiripan karakteristik dari lima variabel. Klaster yang memiliki nilai pusat (*centroid*) positif pada variabel luas panen, produksi, produktivitas, tinggi wilayah, dan curah hujan diinterpretasikan sebagai klaster unggulan yang merepresentasikan wilayah dengan potensi produksi jagung yang tinggi.
- 3 Dengan mengamati konsistensi hasil klasterisasi selama enam tahun yaitu dari tahun 2017 hingga 2022, dapat disimpulkan bahwa terdapat lima wilayah yang secara umum tergolong dalam klaster unggulan, yaitu Kabupaten Pesisir Selatan, Kabupaten Agam, Kabupaten Pasaman, Kabupaten Solok Selatan, dan Kabupaten Pasaman Barat. Wilayah-wilayah ini menunjukkan performa yang relatif stabil dan unggul dalam hal luas panen, tingkat produksi, dan produktivitas, serta didukung oleh kondisi agroklimat yang mendukung, seperti ketinggian dan curah hujan yang sesuai.

## 5. Referensi

- [1] V. N. R. Nani, Y. Boekoesoe, and Y. Bakari, "Analisis Biaya dan Pendapatan Usahatani Jagung di Desa Ayumolingo Kecamatan Pulubala," *AGRINESIA: Jurnal Ilmiah Agribisnis*, vol. 6, no. 2, pp. 157–163, Aug. 2022, doi: 10.37046/agr.v6i2.15917.
- [2] T. N. Rahmah, "Analisis Komparatif Penerapan Biotehnologi pada Padi dan Jagung di Indonesia Periode 2020–2025," *Agrifo : Jurnal Agribisnis Universitas Malikussaleh*, vol. 10, no. 2, pp. 93–101, Oct. 2025, doi: 10.29103/ag.v10i2.24174.
- [3] H. Purwanto and Z. Zulkarnain, "Peran Biotehnologi Probiotik Asal Dadih Sumatera Barat terhadap Performa Produksi Itik Lokal Sumatera Barat," *Zoologi: Jurnal Ilmu Peternakan, Ilmu Perikanan, Ilmu Kedokteran Hewan*, vol. 3, no. 2, pp. 64–72, Jul. 2025, doi: 10.62951/ZOOLOGI.V3I2.204.
- [4] F. Nuraeni and M. H. M. Firdaus, "Model Klasterisasi Data Pertanian Menggunakan Algoritma K-Means," *Jurnal Algoritma*, vol. 22, no. 2, pp. 1–12, Nov. 2025, doi: 10.33364/algoritma/v.22-2.2054.
- [5] N. Mukhlishah, D. Risal, R. Rahmawati, and A. Hafidah, "Penyuluhan Analisis Usaha Tani Jagung Kelurahan Sombalabella, Takalar," *To Maega : Jurnal Pengabdian Masyarakat*, vol. 5, no. 1, pp. 114–122, Jan. 2022, doi: 10.35914/tomaega.v5i1.970.
- [6] D. N. N. Mononimbar, O. E. H. Laoh, and J. S. Tambas, "Analisis Pendapatan Usaha Tani Jagung di Kelurahan Kawangkoan Bawah Kabupaten Minahasa Selatan," *AGRI-SOSIOEKONOMI*, vol. 18, no. 1, pp. 15–22, Apr. 2024, doi: 10.35791/agrsosek.v18i1.55165.
- [7] S. S. Pamusu and Y. Paelo, "Analisis Pendapatan dan Kelayakan Usahatani Jagung Nasa 29 di Kecamatan Pamona Puselemba Kabupaten Poso," *Jurnal Ekonomi Pertanian dan Agribisnis*, vol. 7, no. 1, p. 261, Feb. 2023, doi: 10.21776/ub.jepa.2023.007.01.23.
- [8] N. D. R. Kaban, T. M. Katiandagho, and J. Baroleh, "Analisis Risiko Usaha Tani Jagung di Desa Lompad Baru Kecamatan Ranoyapo Kabupaten Minahasa Selatan," *AGRI-SOSIOEKONOMI*, vol. 19, no. 1, pp. 1–14, Jan. 2023, doi: 10.35791/agrsosek.v19i1.45954.



- [9] Moh. I. A. Rizqoni, A. Y. Arum, and M. Su'udi, "Pemanfaatan Tanaman Produk Rekayasa Genetika di Indonesia," *Jurnal Penelitian Sains*, vol. 26, no. 3, p. 354, Dec. 2024, doi: 10.56064/jps.v26i3.1031.
- [10] A. M. Widayasi, T. Sugiarti, and M. Hayati, "Analisis Perbandingan Pendapatan Usaha Tani Jagung Hibrida Madura-3 Petani Mitra dan Petani Non Mitra di Kab.Pamekasan," *Jurnal Ekonomi Pertanian dan Agribisnis*, vol. 7, no. 2, p. 646, Apr. 2023, doi: 10.21776/ub.jepa.2023.007.02.19.
- [11] M. A. Septianto, A. Faqih, and A. R. Rinaldi, "Klasterisasi Data Produksi Pertanian di Kabupaten Cirebon dengan Algoritma K-Means," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, pp. 1–11, Apr. 2025, doi: 10.23960/jitet.v13i2.6174.
- [12] N. Ahsina, F. Fatimah, and F. Rachmawati, "Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit dengan Menggunakan Metode K-Means Clustering," *Jurnal Ilmiah Teknologi Infomasi Terapan*, vol. 8, no. 3, pp. 1–11, Aug. 2022, doi: 10.33197/jitter.vol8.iss3.2022.883.
- [13] R. R. Ahmed, D. Streimikiene, J. Streimikis, and I. Siksnelyte-Butkiene, "A Comparative Analysis of Multivariate Approaches for Data Analysis in Management Sciences," *E+M Ekonomie a Management*, vol. 27, no. 1, pp. 192–210, Mar. 2024, doi: 10.15240/tul/001/2024-5-001.
- [14] R. Rahmadini, E. E. L. Lubis, A. Priansyah, R. Yolanda, and T. Meutia, "Penerapan Data Mining untuk Memprediksi Harga Bahan Pangan di Indonesia Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Mahasiswa Akuntansi Samudra*, vol. 4, no. 4, pp. 223–235, Aug. 2023, doi: 10.33059/jmas.v4i4.7074.
- [15] D. A. Shafiq, M. Marjani, R. A. A. Habeeb, and D. Asirvatham, "Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review," *IEEE Access*, vol. 10, pp. 72480–72503, 2022, doi: 10.1109/ACCESS.2022.3188767.
- [16] O. Uchendu, K. O. Omomo, and A. E. Esiri, "The Concept of Big Data and Predictive Analytics in Reservoir Engineering: The Future of Dynamic Reservoir Models," *Computer Science & IT Research Journal*, vol. 5, no. 11, pp. 2562–2579, Nov. 2024, doi: 10.51594/csitrj.v5i11.1708.
- [17] E. Teye *et al.*, "Quick Determination of Soil Quality Using Portable Spectroscopy and Efficient Multivariate Techniques," *Journal of Spectroscopy*, vol. 2023, no. 1, pp. 1–14, Jan. 2023, doi: 10.1155/2023/2024318.
- [18] R. H. Pratiwi, "Peranan Biotehnologi dalam Mengatasi Multikrisis," *Faktor Exacta*, vol. 3, no. 2, pp. 158–166, Jul. 2015, doi: 10.30998/FAKTOREXACTA.V3I2.22.
- [19] G. D. Krishnamoorthy and K. Balasubramanian, "Multivariate Technique for The Prediction and Classification of Brain Tumor using Deep Shallow Network," *Appl Soft Comput*, vol. 164, p. 111962, Oct. 2024, doi: 10.1016/j.asoc.2024.111962.