

## KLASTERISASI PROVINSI DI INDONESIA BERDASARKAN PRODUKTIVITAS KOMODITAS PANGAN MENGGUNAKAN ALGORITMA K-MEANS

AdityaNovita <sup>1</sup>, Iin Ernawati S.Kom., M.Si. <sup>2</sup>, Nurul Chamidah S.Kom, M.Kom. <sup>3</sup>,  
S1 Informatika / Fakultas Ilmu Komputer

Program Studi Informatika, Universitas Pembangunan Nasional Veteran Jakarta Jl. RS. Fatmawati Raya, Pd.  
Labu, Kec. Cilandak, Kota Depok, Jawa Barat 12450

email: [adityanov@upnvj.co.id](mailto:adityanov@upnvj.co.id)<sup>1</sup>, [iinernawati@upnvj.ac.id](mailto:iinernawati@upnvj.ac.id)<sup>2</sup>, [nurul.chamidah@upnvj.ac.id](mailto:nurul.chamidah@upnvj.ac.id)<sup>3</sup>

**Abstract** – This research was conducted with the aim of clustering provinces based on harvested area, production, and productivity of food commodities in Indonesia. Data sourced from the website of the Ministry of Agriculture. The data of this research include data on harvested area, production, and productivity of provinces in Indonesia from 2017 to 2019. The study was conducted using K-Means in grouping a data and evaluated by calculating (Sum of Square Error) SSE in order to find the optimal cluster. This research was executed using Google Collaboratory and the language used was python programming. The results of this provincial clustering study resulted in the optimal cluster at  $k=3$  with a difference in SSE value of 241.05797006047. The results of clustering in cluster 0 (medium) amount to 29 data with the characteristics that the province has a more dominant variable whose value is lower than cluster 1 and higher than cluster 2, in cluster 1 (high) there are 64 data characteristic of the province having a more dominant variable whose value is higher. higher than clusters 0 and 2, in cluster 2 (low) there are 9 data with provincial characteristics having more dominant variables whose values are lower than clusters 0 and 1.

**Keywords** – Food, Clustering, K-Means, Elbow

**Abstrak** – Penelitian ini dilakukan bertujuan untuk klasterisasi provinsi berdasarkan luas panen, produksi, dan produktivitas pangan komoditas di Indonesia. Data bersumber dari website Kementerian Pertanian. Data penelitian ini mencakup data luas panen, produksi, dan produktivitas provinsi di Indonesia pada tahun 2017 hingga 2019. Penelitian dilakukan menggunakan K-Means dalam pengelompokan suatu data dan dievaluasi dengan perhitungan (Sum of Square Error) SSE agar ditemukan klaster yang optimal. Penelitian ini dieksekusi menggunakan Google Colaboratory dan bahasa yang digunakan yaitu pemrograman python. Hasil penelitian klasterisasi provinsi ini dihasilkan klaster optimal pada  $k=3$  dengan selisih nilai SSE sebesar 241.05797006047. Hasil klasterisasi pada klaster 0 (sedang) berjumlah 29 data dengan karakteristik yaitu provinsi memiliki variabel lebih dominan yang nilainya lebih rendah dari klaster 1 dan lebih tinggi dari klaster 2, pada klaster 1 (tinggi) berjumlah 64 data karakteristik provinsi memiliki variabel lebih dominan nilainya lebih tinggi dari klaster 0 dan 2, pada klaster 2 (rendah) berjumlah 9 data dengan karakteristik provinsi memiliki variabel lebih dominan nilainya lebih rendah dari klaster 0 dan 1.

**Kata Kunci** – Pangan, Clustering, K-Means, Elbow

### I. PENDAHULUAN

Pangan sangat terus dibutuhkan seiring pertambahan penduduk di Indonesia. Produktivitas suatu pangan dipengaruhi oleh luas panen dan jumlah produksi komoditas pangan. Adanya kecenderungan meluasnyakonversi suatu lahan pertanian ke non pertanian, hal ini menjadi pemicu agar produktivitas pertanian ditingkatkan sebagai strategi untuk peningkatan kapasitas produksi komoditas pangan. Pada penelitian terdahuluyang dilakukan oleh Penelitian yang dilakukan oleh Tendean & Purba bertujuan untuk mengetahui pengelompokan 34 provinsi di Indonesia berdasarkan hasil produksi bahan pangan yang dihasilkan dengan menggunakan algoritma K-Means. Dalam penelitian, data terbagi menjadi 3 kategori harga pangan atau *cluster*. Hasil penelitian pada masing-masing *cluster* yaitu pada cluster 0 (tinggi) terdapat 27 provinsi, *cluster* 1 (sedang) terdapat 4 provinsi, dan *cluster* 2 (rendah) terdapat 3 provinsi [1]. Sedangkan pada penelitian terdahulu yang dilakukan oleh Hartanti (2020) menggunakan metode elbow untuk penentuan jumlah klaster optimum untuk mengukur kesiapan siswa SMK dalam melakukan ujian nasional, dalam penelitian klaster yang terbentuk yaitu 3 klaster atau kategori kesiapan siswa [2].

Pada penelitian yang akan dilakukan ini yaitu melakukan pengelompokan menggunakan algoritma K-Means menggunakan metode elbow terhadap provinsi di Indonesia berdasarkan luas panen, produksi, dan produktivitas pangan komoditas. Penelitian ini dilakukan bertujuan untuk pengelompokan wilayah kota berdasarkan harga pangan pasar tradisional, sehingga luaran yang diperoleh dari hasil penelitian adalah jumlah klaster Provinsi di

Indonesia dari tahun 2017 hingga 2019. Manfaat dari hasil luaran tersebut yaitu penelitian dapat menghasilkan informasi mengenai pengelompokan provinsi di Indonesia dari tahun 2017 hingga 2019 berdasarkan luas panen, produksi, dan produktivitas pangan komoditas.

## II. LANDASAN TEORI

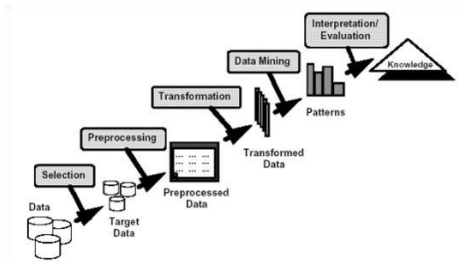
### 2.1. Pangan

Pangan adalah sesuatu yang dihasilkan dari ragam hayati hasil tani, kebun, hutan, ternak, dan perairan yang diolah ataupun tidak hasil olahan yang dikonsumsi manusia sebagai makanan dan minuman, termasuk bahan dasar atau baku, bahan tambahan, dan bahan lain pangan yang dimanfaatkan untuk tahap persiapan, pengerjaan atau pengolahan, dan produksi makanan dan minuman sebagaimana tertulis dalam Undang-Undang Republik Indonesia Nomor 18 Tahun 2012 Pasal 1 Ayat 1.

### 2.2. Data Mining

*Data mining* atau *Knowledge Discovery in Database (KDD)* adalah kegiatan yang tahapnya meliputi pengumpulan suatu data, pengetahuan data dalam jumlah besar, informasi, keteraturan, pola, atau penggunaan data masa lalu untuk menemukan suatu hubungan. Hasil keluaran dari data mining digunakan sebagai salah satu cara menyelesaikan persoalan untuk meningkatkan pengambilan suatu keputusan yang terjadi pada masa depan (Buulolo 2020, p.5). *Data Mining* yaitu tahap menganalisa suatu data yang berjumlah besar yang bertujuan untuk membuat keputusan yang menghasilkan informasi penting dalam data (Arta et al., 2020).

Tahapan KDD menurut Muhammad Iqbal [5]:



Gambar 1. Tahapan KDD [5]

#### 1. Data Selection

Tahap ini diperlukan sebelum tahap penggalian atau pencarian informasi. Penyimpanan hasil pada tahap ini dilakukan untuk melakukan proses data mining ke dalam berkas yang berbeda dengan basis data operasional.

#### 2. Pre-processing / Cleaning

Tahap cleaning sebagai fokus dalam KDD ini, yaitu tahap pembersihan terhadap suatu data agar data dapat diproses dalam data mining.

#### 3. Transformation

Pada tahap transformasi ini yaitu dilakukan tahap coding agar data dalam proses data mining sesuai. Dalam tahap ini coding dilakukan dengan proses yang kreatif dan harus sesuai dengan pola atau jenis informasi dibutuhkan dalam basis data.

#### 4. Data Mining

Data mining yaitu tahap pencarian pola terhadap suatu data agar menghasilkan informasi yang bermanfaat dan menarik menggunakan suatu metode atau teknik tertentu.

#### 5. Interpretation/ Evaluation

Evaluasi ini merupakan tahap pengecekan terhadap informasi yang diperoleh berhubungan atau tidak sesuai dengan dugaan sementara sebelumnya.

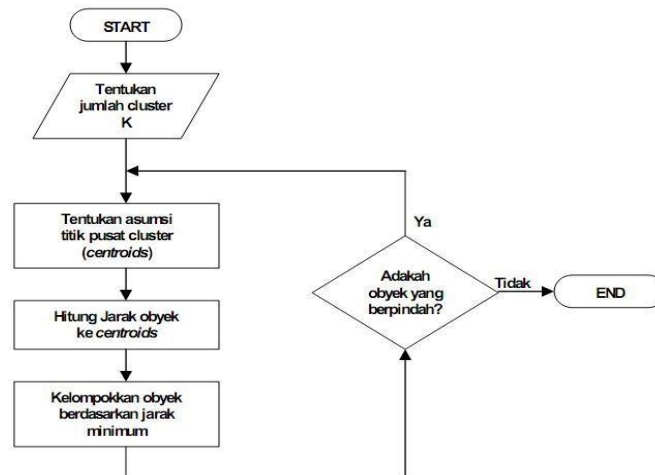
### 2.3. Clustering

*Clustering* yaitu proses yang dilakukan dengan melakukan pengelompokan terhadap objek menurut informasi yg diperoleh berdasarkan data yg mengungkapkan interaksi antar objek menggunakan prinsip agar memaksimalkan komponen pada satu grup dan meminimalkan kecenderungan antar kluster [6]. *Clustering* dapat mengetahui pola distribusi secara menyeluruh, dan mendapatkan informasi hubungan menarik pada indikator data. Data mining berfokus pada bagaimana mendeteksi *cluster* dalam database besar secara efektif dan efisien [7].

### 2.4. K-Means Clustering

K-Means adalah teknik pengelompokan non-hierarki yang bersifat unsupervised. Definisi K-Means yaitu teknik pengelompokan berdasarkan jarak dengan membagi suatu data menjadi kelompok sehingga data yang serupa masuk satu klaster dan jika tidak mirip akan masuk ke dalam klaster lain, dan metode ini hanya untuk atribut berupa numerik [8].

Berikut flowchart langkah-langkah metode K-Means sebagai berikut [9]:



Gambar 2. Flowchart metode K-Means

- Pada proses pembagian data, harus ditentukannya jumlah *cluster*.
- Pada pemilihan awal pusat klaster, centroid awal sebagai titik pusat cluster yang pemilihannya secara acak.
- Temukan pada semua data yang memiliki jarak paling dekat dari setiap data terhadap centroid dalam proses pengolahan suatu data menggunakan perhitungan jarak euclidean.

Berikut merupakan persamaan Euclidian Distance:

$$d(x_q, \mu_k) = \sqrt{\sum (x_q - \mu_k)^2} \quad (1)$$

Dimana :

$x_q$ : objek data

$\mu_k$ : centroid

- Jarak pada titik pusat centroid terdekat itu dikelompokkan setiap data, selanjutnya ditemukan pusat klaster yang baru setelah anggota klaster baru terbentuk.
- Nilai rata-rata pusat cluster merupakan parameter yang digunakan, kalau anggota tiap cluster tidak ada yang berubah maka iterasi berhenti atau selesai.
- Lakukan langkah c, iterasi akan berhenti hingga pusat klaster atau centroid sudah tidak ada lagi yang berubah atau tetap.

Karakteristik *K-means* [8]:

- K-means* pada tahap pengelompokan sangat efisien.
- K-means* bersifat sensitif pada pemilihan awal titik pusat pada klaster yang dipilih dengan acak.
- Memungkinkan pada suatu kelompok ada yang tidak memiliki suatu objek atau anggota dalam klaster.
- Pada hasil pengelompokan dengan *K-means* memiliki sifat yang tidak tetap atau selalu berubah-ubah.

## 2.5. Metode *Elbow*

Dalam menentukan jumlah optimal klaster yaitu metode *elbow* digunakan untuk membandingkan antara jumlah cluster yang terbentuk dari perhitungan *Sum of Square Error (SSE)* setiap nilai *cluster* yang membentuk siku dalam suatu titik pada grafik [2]. Rumus SSE sebagai berikut:

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|x_i - c_k\|_2^2 \quad (2)$$

Dimana :

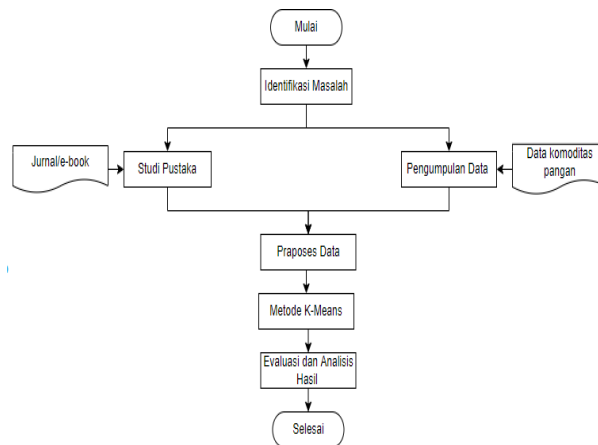
$k$  = jumlah nilai kluster  
 $x_i$  = data ke  $-i$   
 $C_k$  = centroid pada *cluster*

## 2.6. Python

Python merupakan bahasa pemrograman yang banyak digunakan dalam membuat suatu program. Sintaks bahasa pemrograman Python tidak terlalu sulit, maka Python merupakan bahasa pemrograman tingkat tinggi yang mudah dalam pengimplementasiannya. Saat membuat kode program dalam bahasa pemrograman Python harus mengikuti aturan dalam pembuatan program [10].

## III. METODE PENELITIAN

Berikut tahapan penelitian pada gambar 1 :



Gambar 3. Flowchart tahapan penelitian

### 3.1. Identifikasi Masalah

Pada tahap ini merupakan tahap awal penelitian yang akan dilakukan untuk mencari fakta-fakta adanya permasalahan dan menjelaskan masalah yang terjadi agar masalah tersebut ditemukan solusinya. Dalam penelitian ini untuk menyelesaikan permasalahan dilakukan klusterisasi untuk mengetahui jumlah kluster pada provinsi di Indonesia. Terkait permasalahan tersebut maka digunakan data bersumber website Kementerian Pertanian yang terdiri dari data luas panen, produksi, dan produktivitas pangan komoditas provinsi di Indonesia pada tahun 2017, 2018, dan 2019.

### 3.2. Studi Pustaka

Tahap ini melakukan studi literatur agar dalam penelitian ini diperoleh bahan kajian pustaka dari penelitian terdahulu seperti bersumber dari jurnal dan e-book yang berkaitan dengan penelitian ini.

### 3.3. Pengumpulan Data

Tahap ini dilakukan eksplorasi data yang berkaitan dengan permasalahan yang terjadi agar mendapat informasi untuk mencapai tujuan dalam penelitian. Dalam pengumpulan data, data yang digunakan harus benar-benar atau berguna agar dapat diolah. Data dalam penelitian menggunakan data website Kementerian Pertanian yaitu [www.pertanian.go.id](http://www.pertanian.go.id) dengan link <https://www.pertanian.go.id/home/?show=page&act=view&id=61> yang terdiri dari data luas panen, produksi, dan produktivitas . Data pangan komoditas dalam penelitian ini yang digunakan yaitu 7 komoditas diantaranya bawang merah, cabe besar, tomat, jahe, kencur, kunyit, dan lengkuas selama tiga tahun dari tahun 2017 hingga 2019 pada Provinsi di Indonesia. Pada website kementerian pertanian hanya tersedia data selama tiga tahun pada komoditas pangan tersebut.

### 3.4. Praproses Data

Tahap praproses yang dilakukan dalam penelitian ini yaitu pada tahap awal melakukan *cleaning* pada data yang nilainya tidak ada atau bertanda (-) karena data tidak tersedia pada microsoft excel . Data yang bertanda (-) merupakan data yang tidak menghasilkan produksi komoditas pangan. Tahap *cleaning* yang dilakukan yaitu pada data yang kosong atau bertanda (-) dilakukan pengubahan tanda (-) ke nilai 0. Tahap *cleaning* dilakukan agar data dapat dilakukan normalisasi dengan baik.

Tahap praproses yang dilakukan selanjutnya setelah melakukan *cleaning* data yaitu tahap normalisasi pada data yang gunanya agar datanya seimbang jika dilakukan perhitungan. Normalisasi data pada penelitian ini yaitu menggunakan *z-score normalization*.

### 3.5. Metode K-Means

Pada penelitian ini menggunakan bahasa pemrograman python dalam mengimplementasikan metode K-Means. Implementasikan pengelompokan data menggunakan persamaan jarak euclidean dengan uji *cluster* uji k=2 hingga k=6. Tahap ini, pusat klaster awal pada data ditentukan secara acak. *K-Means* kemudian menguji setiap anggota dari suatu data dan menandai anggota tersebut dengan salah satu titik centroid pusat cluster yang ditentukan sesuai dengan jarak terdekat antara anggota dengan setiap cluster. Semua anggota pada data digolongkan pada setiap pusat *cluster* dan titik pusat *cluster* dihitung kembali sampai akhirnya terbentuk titik pusat *cluster* baru. *K-Means* dengan Persamaan jarak *euclidean* dalam penelitian ini untuk mengelompokkan provinsi di Indonesia pada tahun 2017,2018, dan 2019 berdasarkan luas panen, produksi, dan produktivitas pangan komoditas. Pengelompokan provinsi berdasarkan produktivitas digunakan tiga indikator pada masing-masing variabel karena produktivitas komoditas pangan dipengaruhi jumlah produksi dan luas panen.

### 3.6. Evaluasi dan Analisis Hasil

Pada penelitian ini tahap evaluasi dengan menggunakan metode *elbow*. Metode *elbow* untuk penentuan cluster optimal dengan persentase perbandingan antara klaster sebelum dan sesudah yang mengalami penurunan terbesar pada grafik di suatu titik yang membentuk suatu siku. Dari hasil persentase yang dihitung dijadikan perbandingan antara penjumlahan jumlah klaster. *Sum of Square Error* untuk evaluasi jumlah klaster dari hasil pengujian K-Means. *SSE* dihitung menggunakan rumus persamaan berikut:

$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|x_i - C_k\|_2^2$$

Dimana :

k = jumlah nilai kluster

xi = data ke -i

Ck = centroid pada cluster

Setelah didapat klaster terbaik menggunakan metode *elbow*, kemudian menentukan nama kategori pada masing-masing label atau klaster. Dalam penentuan kategori diperlukan analisis perbandingan nilai rata-rata antar klaster.

## IV. HASIL DAN PEMBAHASAN

### 4.1. Pengumpulan Data

Data dalam penelitian menggunakan data bersumber dari website Kementerian Pertanian yang terdiri dari data Luas Panen, Produksi, dan Produktivitas komoditas pangan pada tahun 2017 hingga 2019 provinsi di Indonesia. Komoditas diantaranya yaitu bawang merah, cabe besar, tomat, jahe, kencur, kunyit, dan lengkuas. Total atribut pada data berjumlah 23 atribut. Atribut yang dilakukan perhitungan menggunakan clustering K-Means berjumlah 21 atribut yaitu indikator luas panen, produksi, dan produktivitas pada masing-masing komoditas. Atribut pendukung pada penelitian ini yaitu tahun dan nama provinsi. Dalam penelitian ini hanya menggunakan 7 komoditas pangan tersebut karena penelitian ini terfokus pada bumbu pelengkap masakan.

Pengumpulan data pada penelitian ini yaitu data diperoleh dari website kementerian pertanian dengan link <https://www.pertanian.go.id/home/?show=page&act=view&id=61>. Data yang tersedia pada website kementerian pertanian yaitu data komoditas pangan dengan indikator luas panen, produksi, dan produktivitas. Data pada website kementerian pertanian yaitu file masing-masing komoditas dan antara indikator satu dengan yang lainnya terpisahkan dengan komoditas pangan lainnya, maka penelitian ini dilakukan penggabungan data.

### 4.2. Hasil Klasterisasi *Clustering K-Means*

Pada penelitian ini menggunakan bahasa pemrograman python dalam mengimplementasikan metode K-Means. Implementasikan pengelompokan data menggunakan persamaan jarak euclidean dengan uji *cluster* uji k=2 hingga k=6. Berikut merupakan hasil klaster pada uji k=3 :

**Tabel 1. Hasil Klasterisasi dari tahun 2017 hingga 2019**

Provinsi	2017	2018	2019
Aceh	Klaster 0 (Sedang)	Klaster 0 (Sedang)	Klaster 0 (Sedang)
Sumatera Utara	Klaster 0 (Sedang)	Klaster 0 (Sedang)	Klaster 0 (Sedang)
Sumatera Barat	Klaster 0 (Sedang)	Klaster 0 (Sedang)	Klaster 0 (Sedang)
Riau	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Jambi	Klaster 0 (Sedang)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Sumatera Selatan	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Bengkulu	Klaster 0 (Sedang)	Klaster 0 (Sedang)	Klaster 0 (Sedang)
Lampung	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)

Kepulauan Bangka Belitung	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Kepulauan Riau	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
DKI Jakarta	Klaster 0 (Sedang)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Jawa Barat	Klaster 1 (Tinggi)	Klaster 1 (Tinggi)	Klaster 1 (Tinggi)
Jawa Tengah	Klaster 1 (Tinggi)	Klaster 1 (Tinggi)	Klaster 1 (Tinggi)
DI Yogyakarta	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Jawa Timur	Klaster 1 (Tinggi)	Klaster 1 (Tinggi)	Klaster 1 (Tinggi)
Banten	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Bali	Klaster 0 (Sedang)	Klaster 0 (Sedang)	Klaster 0 (Sedang)
Nusa Tenggara Barat	Klaster 0 (Sedang)	Klaster 0 (Sedang)	Klaster 0 (Sedang)
Nusa Tenggara Timur	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Kalimantan Barat	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Kalimantan Tengah	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Kalimantan Selatan	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Kalimantan Timur	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Kalimantan Utara	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Sulawesi Utara	Klaster 2 (Rendah)	Klaster 0 (Sedang)	Klaster 2 (Rendah)
Sulawesi Tengah	Klaster 2 (Rendah)	Klaster 0 (Sedang)	Klaster 2 (Rendah)
Sulawesi Selatan	Klaster 0 (Sedang)	Klaster 0 (Sedang)	Klaster 0 (Sedang)
Sulawesi Tenggara	Klaster 2 (Rendah)	Klaster 0 (Sedang)	Klaster 2 (Rendah)
Gorontalo	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Sulawesi Barat	Klaster 0 (Sedang)	Klaster 0 (Sedang)	Klaster 0 (Sedang)
Maluku	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Maluku Utara	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 0 (Sedang)
Papua Barat	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)
Papua	Klaster 2 (Rendah)	Klaster 2 (Rendah)	Klaster 2 (Rendah)

Berikut merupakan visualisasi Peta hasil klasterisasi Provinsi komoditas pangan pada tahun 2017 :



**Gambar 4. Peta hasil klasterisasi provinsi tahun 2017**

Berikut merupakan visualisasi hasil klasterisasi provinsi pada tahun 2018:



**Gambar 5. Peta hasil klasterisasi provinsi tahun 2018**

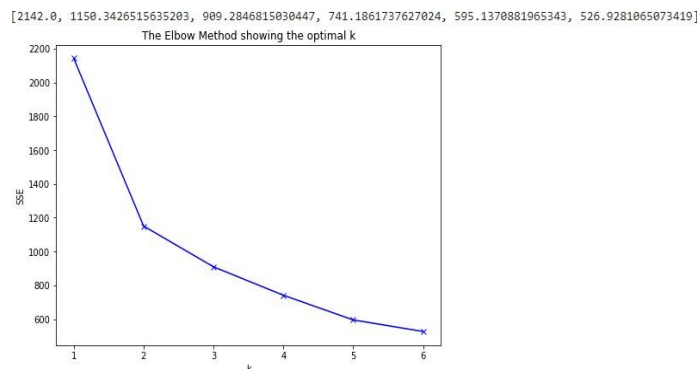
Berikut merupakan visualisasi hasil klasterisasi provinsi pada tahun 2019:



**Gambar 6. Peta hasil klasterisasi provinsi tahun 2019**

#### 4.3. Evaluasi dan Analisis Hasil

Dalam klasterisasi dilakukan uji metode elbow untuk menentukan jumlah kluster optimal dengan membandingkan nilai SSE antar kluster dan membentuk suatu siku pada grafik elbow di suatu titik. Metode elbow dilakukan setelah dilakukan uji metode K-Means Clustering untuk mendapatkan total nilai SSE pada uji kluster  $k=2$  hingga  $k=6$  pada data yang diuji. Dalam uji metode elbow jika nilai suatu  $k$  semakin besar maka nilai SSE yang dihasilkan akan semakin kecil. Berikut merupakan grafik metode elbow pada 102 data sebagai berikut:



**Gambar 7. Grafik Elbow pada 102 data**

Berikut merupakan hasil pengujian metode elbow pada 102 data. Pada metode ini uji  $k$  dimulai dari  $k=2$  hingga  $k=6$ . Berikut pada Tabel 5. merupakan hasil pengujian perhitungan nilai SSE dan selisih nilai SSE antar kluster sebelum dan sesudah pada 102 data.

**Tabel 2. Hasil perhitungan metode elbow pada 102 data**

Klaster	Nilai SSE	Selisih Nilai SSE
K2	1150.3426515635203	1150.3426515635203
K3	909.2846815030447	241.05797006047
K4	741.1861737627024	168.09850774034
K5	595.1370881965343	146.04908556616
K6	526.9281065073419	68.208981612311

Pada Tabel 5. uji coba metode elbow 102 data bahwa  $k=2$  memiliki nilai SSE 1150.3426515635203,  $k=3$  sebesar 909.2846815030447,  $k=4$  sebesar 741.1861737627024,  $k=5$  sebesar 595.1370881965343, dan  $k=6$  sebesar 526.9281065073419. Dari hasil perhitungan selisih nilai SSE,

didapat kluster optimal yaitu pada kluster 3 karena terjadi penurunan nilai terbesar dari kluster 2 ke kluster 3 dengan selisih nilai sebesar 241.05797006047. Berikut merupakan hasil output jumlah data pada masing-masing kluster :

**Tabel 3. Jumlah data provinsi setiap kluster**

Kluster	Total data
0	29
1	9
2	64

4.3.1. Karakteristik *cluster*

Berikut merupakan analisis kluster atau karakteristik setiap kluster hasil klusterisasi provinsi berdasarkan luas panen, produksi, dan produktivitas komoditas pangan:

1. Klaster 0

Klaster 0 merupakan kluster yang potensi pertaniannya dengan nama kategori “**Sedang**”. Pada klaster 0, provinsi yang masuk ke dalam kluster ini memiliki karakteristik dengan variabel komoditas pangan lebih dominan memiliki nilai rata-rata yang lebih rendah dari klaster 1 dan lebih tinggi dari klaster 2. Berikut merupakan rentang nilai pada klaster 0 pada setiap indikator:

**Tabel 4. Rentang nilai setiap atribut pada klaster 0**

kluster	Luas Panen Bawang Merah	Produksi Bawang Merah	Produktivitas Bawang Merah	Luas Panen Cabe Besar	Produksi Cabe Besar	Produktivitas Cabe Besar	Luas Panen Tomat
0	0 -19341	0 - 212885	0 -14,97	0 - 16410	0 - 159131	0 - 15,26	0 - 5525

Produksi Tomat	Produktivitas Tomat	Luas Panen Jahe	Produksi Jahe	Produktivitas Jahe	Luas Panen Kencur
0 - 146829	0 - 36,6	1949 - 3921274	28413 - 14025136	1,82 - 14,11	743 - 519011

Produksi Kencur	Produktivitas Kencur	Luas Panen Kunyit	Produksi Kunyit	Produktivitas Kunyit	Luas Panen Lengkuas	Produksi Lengkuas	Produktivitas Lengkuas
5037 - 1133498	1,02 - 7,49	1376 - 4020667	16803 - 12335424	1,7 - 11,77	1177 - 836440	15164 - 3133000	1,91 - 12,46

2. Klaster 1

Klaster 1 merupakan kluster yang potensi pertaniannya dengan nama kategori “**Tinggi**”. Pada klaster 1, provinsi yang masuk ke dalam kluster ini memiliki karakteristik variabel komoditas pangan lebih dominan memiliki rata-rata luas panen yang lebih luas dari klaster 0 dan 2, jumlah produksi yang lebih banyak dari klaster 0 dan 2, produktivitas lebih tinggi dari klaster 0 dan 2.

**Tabel 5. Rentang nilai setiap atribut pada klaster 1**

klaster	Luas Panen Bawang Merah	Produksi Bawang Merah	Produktivitas Bawang Merah	Luas Panen Cabe Besar	Produksi Cabe Besar	Produktivitas Cabe Besar
1	15404 - 51155	166865 - 481890	8,24 - 11,04	12190 - 28481	91965 - 274311	6,86 - 13,67

Luas Panen Tomat	Produksi Tomat	Produktivitas Tomat	Luas Panen Jahe	Produksi Jahe	Produktivitas Jahe	Luas Panen Kencur
------------------	----------------	---------------------	-----------------	---------------	--------------------	-------------------



4000 - 10171	65585 - 295321	15,31 - 29,71	10868882 - 49125970	26966783 - 77241049	1,44 - 2,48	2393065 - 7549232
--------------	----------------	---------------	---------------------	---------------------	-------------	-------------------

Produksi Kencur	Produktivitas Kencur	Luas Panen Kunyit	Produksi Kunyit	Produktivitas Kunyit	Luas Panen Lengkuas	Produksi Lengkuas	Produktivitas Lengkuas
3802062 - 8854369	1,16 - 1,91	3167957 - 43621260	7828267 - 117108216	1,46 - 3,05	3185116 - 6820322	10091992 - 18765630	1,99 - 3,36

### 3. Klaster 2

Klaster 2 merupakan klaster yang potensi pertaniannya dengan nama kategori “Rendah”. Pada klaster 2 provinsi yang masuk ke dalam klaster ini memiliki karakteristik yaitu dominasi variabel paling banyak memiliki nilai yang lebih rendah dari klaster 0 dan 1. variabel komoditas pangan rata-rata memiliki luas panen yang lebih kecil, jumlah produksi yang lebih sedikit, produktivitas lebih rendah dari klaster 0 dan 1. Berikut merupakan rentang nilai setiap atribut pada klaster 2 :

**Tabel 6. Rentang nilai setiap atribut pada klaster 2**

kluster	Luas Panen Bawang Merah	Produksi Bawang Merah	Produktivitas Bawang Merah	Luas Panen Cabe Besar	Produksi Cabe Besar	Produktivitas Cabe Besar	Luas Panen Tomat
2	0 - 1740	0 - 16999	0 - 9,97	0 - 6886	0 - 50203	0 - 11,79	0 - 4158

Produksi Tomat	Produktivitas Tomat	Luas Panen Jahe	Produksi Jahe	Produktivitas Jahe	Luas Panen Kencur	Produksi Kencur	Produktivitas Kencur
0 - 42392	0 - 15,36	2444 - 2495316	1550 - 8545276	0,63 - 7,92	0 - 5231934	0 - 7500387	0 - 3,21

Luas Panen Kunyit	Produksi Kunyit	Produktivitas Kunyit	Luas Panen Lengkuas	Produksi Lengkuas	Produktivitas Lengkuas
547 - 2141756	722 - 3708139	0,98 - 6,52	4 - 1009532	2 - 3641410	0,43 - 6,65

## V. HASIL DAN PEMBAHASAN

### 5.1. Kesimpulan

1. Hasil klasterisasi provinsi berdasarkan luas panen, produksi, dan produktivitas pada penelitian ini yaitu hasil klasterisasi pada 102 data, dihasilkan klaster 0 dengan potensi pertanian kategori sedang berjumlah 29 data, klaster 1 dengan potensi pertanian kategori tinggi berjumlah 9 data, dan cluster 2 dengan potensi pertanian kategori rendah berjumlah 64 data. Karakteristik pada klaster 0 (sedang) yaitu provinsi yang masuk ke dalam klaster 0 yaitu yang variabel komoditas pangannya rata-rata dominan lebih tinggi dari klaster 2 tetapi lebih kecil dari klaster 1. Karakteristik pada klaster 1 yaitu provinsi yang masuk ke dalam klaster 1 variabel komoditas pangannya rata-rata dominan lebih tinggi dari klaster 0 dan 2. Karakteristik pada klaster 2 yaitu provinsi yang masuk ke dalam klaster 2 variabel komoditas pangan dominan lebih rendah dari klaster 0 dan 1.

2. Klasterisasi provinsi pada tahun 2017 hingga 2019 pada penelitian ini menggunakan metode *K-Means Clustering* dan evaluasi uji metode elbow. Dari hasil evaluasi dengan metode elbow dihasilkan grafik yang nilai SSE mengalami penurunan terbesar dari k=2 ke k=3 dengan selisih sebesar 241.05797006047, maka k=3 merupakan klaster optimal.

### 5.2. Saran

1. Memperbanyak jumlah data penelitian.
2. Menggunakan metode lain untuk evaluasi seperti silhouette coefficient dan Davies bouldin Index.
3. Menambahkan variabel lainnya seperti jumlah petani agar dapat mengetahui peningkatan potensi pertanian jika dilihat dari banyak petani yang bekerja.

## REFERENSI

- [1] T. Tendean and W. Purba, "Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means," *J. Sains Dan Teknol.*, vol. 1, no. 2, pp. 5–11, 2020.
- [2] N. T. Hartanti, "Jurnal Nasional Teknologi dan Sistem Informasi Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional," vol. 02, pp. 82–89, 2020.
- [3] E. Baulolo, *Data Mining Untuk Perguruan Tinggi*. 2020.
- [4] J. Arta, G. Indrawan, G. R. Dantes, P. Studi, and I. Komputer, "DATA MINING REKOMENDASI CALON MAHASISWA BERPRESTASI DI STMIK DENPASAR MENGGUNAKAN METODE TECHNIQUE FOR OTHERS REFERENCE BY SIMILARITY TO IDEAL SOLUTION."
- [5] M. Iqbal, "KLASTERISASI DATA JAMAAH UMROH PADA AULIYA TOUR & TRAVEL MENGGUNAKAN METODE K-MEANS CLUSTERING PENDAHULUAN Perkembangan teknologi saat ini semakin pesat dan tidak terbendung . Semakin mudahnya mendapatkan informasi dan proses transaksi data yang cepat men," vol. V, no. 2, pp. 97–104, 2019.
- [6] D. D. Darmansah and N. W. Wardani, "Analisis Pesebaran Penularan Virus Corona di Provinsi Jawa Tengah Menggunakan Metode K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 105–117, 2021, doi: 10.35957/jatisi.v8i1.590.
- [7] Y. Darmi and A. Setiawan, "PENERAPAN METODE CLUSTERING K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK," vol. 12, no. 2, pp. 148–157, 2016.
- [8] Suhartini and R. Yuliani, "Infotek : Jurnal Informatika dan Teknologi Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K- Means di Dusun Bagik Endep Sukumulia Timur Infotek : Jurnal Informatika dan Teknologi Pendahuluan masalah kemiskinan belum bis," vol. 4, no. 1, pp. 39–50, 2021.
- [9] E. Muningsih, "KOMPARASI METODE CLUSTERING K-MEANS DAN K-MEDOIDS DENGAN MODEL FUZZY RFM UNTUK PENGELOMPOKAN PELANGGAN," vol. 6, no. 2, 2018.
- [10] I. W. S. Wijaya, I. G. H. W. KS., I. D. M. A. P. S. Bintara, and I. K. G. R. A. Permana, "Program Menghitung Banyak Bata pada Ruangan Menggunakan Bahasa Python," *TIERS Inf. Technol. J.*, vol. 2, no. 1, pp. 12–22, 2021, [Online]. Available: <http://journal.undiknas.ac.id/index.php/tiers/article/view/2840>.