

Optimasi K-Klasterisasi Ketahanan Pangan Kabupaten Jember Menggunakan Metode Elbow

Aski Widdatul Fuadah*, Fajrin Nurman Arifin**, Oktalia Juwita***

Information System, University of Jember

*askiwidda21@gmail.com, **fajrin.pssi@unej.ac.id, ***oktalia@unej.ac.id

Abstrak

Clustering is a process of grouping data based on similarities or similarities with other members in a group. Food security is the condition of a country to provide food for individuals, which does not conflict with beliefs, religion and culture and leads a healthy, active and productive life. Food instability and food insecurity can be caused by many factors, one of which is natural disasters. In 2020, Jember Regency experienced 121 natural disasters. Determination of the optimal K value is done to get the right number of group divisions from the clustering process, in this case using the elbow method. The data used in the clustering process are sub-districts in Jember Regency using transient attributes or natural disaster events. Based on the results of sub-district data grouping from the number of clusters $k=1$ to $k=10$, the optimal k value was found at the value of $k = 4$ with the SSE (Sum of Square Error) value = 24,809.

Kata Kunci: Clustering, Transient, Elbow, Sum Square Error

1. Pendahuluan

Klasterisasi adalah suatu proses pengelompokan data berdasarkan persamaan atau kemiripan dengan anggota lainnya di dalam sebuah kelompok (Tan. et.al., 2004). Dalam proses klasterisasi terdapat beberapa metode yang dapat dipertimbangkan dalam penentuan jumlah kluster. Salah satu metode yang dapat digunakan dalam pencarian kluster optimum adalah metode Elbow (Putu dkk., 2016). Metode elbow adalah suatu metode yang digunakan dalam penentuan jumlah k optimal dari pembentukan sebuah cluster. Berdasarkan penjelasan sebelumnya dalam penentuan jumlah k atau pencarian jumlah k optimum dapat mempertimbangkan beberapa metode. Ananda dan Yamani dalam penelitiannya yang berjudul "Penentuan centroid awal k-means pada proses clustering data evaluasi" menunjukkan bahwa dari perbandingan beberapa metode yang digunakan menunjukkan bahwa elbow merupakan salah satu metode evaluasi yang dapat menghasilkan nilai k optimum terbaik dengan nilai index *silhouette* 0,6772. Beberapa metode yang digunakan dalam penelitian tersebut adalah *Silhouette Density Canopy (SDC)*, *Density Canopy (DC)*, *Silhouette (S)*, *Elbow (E)* dan *Bayesian Information Criterion (BIC)* (Ananda dan Yamani, 2021).

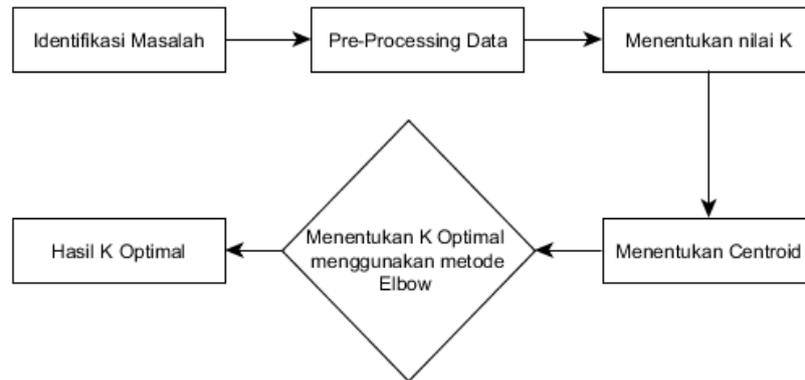
Berdasarkan undang undang nomor 18 tahun 2012 ketahanan pangan adalah kondisi suatu negara untuk menyediakan pangan bagi individu, yang tidak bertentangan dengan keyakinan, agama dan budaya serta menjalani kehidupan yang sehat, aktif dan produktif. Lain halnya dengan ketahanan pangan, kondisi kerawanan pangan dapat terjadi karena disebabkan oleh beberapa faktor. Salah satu faktor yang menyebabkan kondisi kerawanan pangan adalah terjadinya bencana alam yang selanjutnya dalam penelitian ini menjadi variabel atau atribut yang dianalisis menggunakan proses klasterisasi. Petani sebagai salah bagian penting dalam penyediaan bahan makanan sangat terpengaruh oleh kejadian bencana alam pada wilayah atau lahan pertaniannya. Hal tersebut disebabkan karena terjadinya bencana alam yang berdampak pada tidak produktifnya lahan pertanian dari para petani (Oktari, 2018). Hal lainnya diungkapkan bahwa dampak bencana alam sangat berpengaruh pada lahan pertanian sehingga menyebabkan penurunan jumlah produksi (Hapsari dan Rudiarto, 2017). Terdapat beberapa dampak lainnya yang dialami pada suatu wilayah yang terdampak bencana alam. Dampak tersebut diantaranya adalah berkenaan dengan akses pasar, perdagangan, ketersediaan pangan, penurunan pendapatan, kenaikan harga pangan penurunan pendapatan usaha tani dan lapangan kerja (Tirivangasi, 2018). Kabupaten Jember pada tahun 2020 berdasarkan data yang tercatat di BPBD mengalami beberapa kejadian bencana alam. Kejadian tersebut diantaranya adalah 121 kejadian bencana alam, dengan rincian 17 banjir, 16 tanah longsor, 57 angin kencang, 26 kebakaran, 1 erupsi gunung api, 2 kekeringan, dan 1 gempa bumi dengan 2324 KK yang terdampak (Infografis Kejadian Bencana Kabupaten Jember Tahun 2020).

Pada penelitian ini nantinya akan dilakukan analisis untuk mengetahui dan memperoleh nilai k optimum dengan menggunakan metode Elbow. Diharapkan dengan menggunakan metode Elbow nantinya dapat menghasilkan pengelompokan atau proses klasterisasi yang optimal. Dalam penelitian ini untuk

mengetahui jumlah k optimum pada tahapan uji nilai k akan menggunakan nilai $k=1$ hingga $k=10$. Hal tersebut untuk memperoleh nilai k optimal dalam proses klasterisasi data ketahanan pangan di Kabupaten Jember.

2. Metodologi Penelitian

Metodologi penelitian merupakan penjelasan tahapan-tahapan yang dilakukan dalam proses penelitian yang menjadi peta pandu dalam pelaksanaan penelitian. Dalam gambar 1 dijelaskan tahapan yang dilakukan dalam penelitian mulai dari identifikasi permasalahan hingga nilai k telah ditemukan.



Gambar 2.1. Tahapan Penelitian

2.1 Identifikasi Masalah

Tahapan ini merupakan penjelasan tentang permasalahan yang diperoleh dalam objek dan pemilihan solusi yang dapat digunakan dalam penyelesaian masalah tersebut. Dalam hal ini proses identifikasi yang dilakukan diantaranya adalah analisis terhadap atribut-atribut yang digunakan dalam penelitian berdasarkan atribut FSVA (*Food Security and Vulnerability Atlas*).

2.2 Pre-Processing Data

Berdasarkan hasil yang diperoleh pada proses identifikasi masalah selanjutnya dilakukan proses penyiapan data atau pre processing data. Dalam hal ini proses yang dilakukan diantaranya adalah standarisasi data atau penyesuaian data pada masing-masing atribut agar memiliki jenis (jenis dan satuan) data yang sama. Selanjutnya data yang telah mengalami proses standarisasi dilakukan proses normalisasi agar nilai data pada masing-masing atribut seimbang dan tidak mendominasi terhadap atribut lainnya (Mulyana Nasution, 2019).

2.3 Menentukan nilai K

Pada proses ini merupakan proses yang dilakukan dalam bentuk perulangan dari nilai k untuk menentukan nilai k optimal. Nilai K yang digunakan dalam proses ini adalah nilai $k=1$ hingga $k=10$. Jika nilai k belum menunjukkan elbow berdasarkan nilai SSE yang diperoleh, maka proses perulangan akan terus dilakukan hingga $k=10$ atau nilai elbow telah ditemukan.

2.4 Menentukan Centroid

Penentuan nilai centroid dilakukan untuk mendapatkan estimasi nilai centroid awal sampai diperoleh nilai centroid utama atau nilai centroid tidak lagi mengalami perubahan. Dalam proses ini perubahan atau penentuan nilai centroid berdasarkan perbandingan nilai inter dan intra cluster menggunakan persamaan Euclidean Distance. Jika nilai distance intra cluster $>$ inter cluster, maka nilai centroid mengalami perubahan hingga memenuhi batasan nilai intra cluster $<$ inter cluster.

2.5 Metode Elbow

Metode Elbow adalah salah satu metode yang merupakan proses evaluasi sekaligus juga digunakan untuk menemukan nilai k optimum dari perulangan proses uji nilai k . Pencarian nilai k optimum dilakukan dengan melakukan perbandingan nilai SSE (Sum of Square Error) yang disajikan dalam bentuk grafik. Nilai k optimum akan ditemukan saat kondisi Elbow (patahan) telah ditemukan pada grafik tersebut (Yana dkk., 2018).

2.6 Hasil K Optimal

Bagian ini adalah proses terakhir dalam penelitian, proses ini merupakan penjelasan bahwa nilai k optimal telah tercapai dan proses optimasi klasterisasi telah selesai. Proses tersebut dapat dicapai jika nilai k optimal telah ditemukan atau kondisi elbow telah diperoleh. Jika kondisi nilai k optimal belum ditemukan atau kondisi elbow belum diperoleh maka proses penentuan nilai k dikembalikan pada tahapan penentuan nilai k selanjutnya.

3. Hasil dan Pembahasan

3.1. Identifikasi Masalah

Dalam penanganan kondisi kerawanan pangan yang sering terjadi pada wilayah-wilayah NKRI pemerintah Indonesia telah menyusun sebuah pedoman. Pedoman tersebut adalah membuat *Food Security and Vulnerability Atlas* (FSVA). FSVA dalam penentuan kondisi kerawanan pangan suatu wilayah menggunakan 2 indikator kerawanan pangan, indikator tersebut adalah indikator kronis dan transien. Salah satu indikator kerawanan pangan yang dijelaskan dalam FSVA adalah indikator transien yang merupakan kondisi banyaknya atau jumlah bencana alam yang terjadi karena iklim, perubahan cuaca hujan bulanan atau variabilitas curah hujan, rata-rata kehilangan produksi tanaman pangan akibat dampak perubahan iklim (banjir, kebakaran hutan, kekeringan dan badai) dan akibat Organisme Pengganggu Tanaman (OPT) (Pangan dan Pertanian, 2020).

Berdasarkan tersebut penelitian yang dilakukan adalah proses analisis dan optimasi klusterisasi data ketahanan pangan Kabupaten Jember untuk mendapatkan nilai k optimum menggunakan metode Elbow. Selanjutnya untuk objek penelitian yang dilakukan adalah kecamatan-kecamatan yang termasuk dalam wilayah Kabupaten Jember. Pemilihan objek dipengaruhi oleh banyaknya kejadian bencana alam pada Kabupaten Jember di tahun 2020. Selain itu berdasarkan data yang diperoleh menunjukkan bahwa Kabupaten Jember dalam produksi bahan pangannya sangat dipengaruhi kondisi-kondisi yang termasuk dalam indikator transien.

3.2. Pre-Processing Data

Dalam proses klusterisasi sebelum data diolah dan dikelompokkan maka dibutuhkan sebuah proses atau tahapan untuk menyiapkan atau menyajikan data olah. Tahapan tersebut adalah tahap *pre-Processing* data. Pre-Processing data adalah Teknik yang termasuk dalam tahap awal pengolahan yang dilakukan untuk mengolah bahan mentah (*raw data*) yang selanjutnya setelah melalui proses penyiapan atau pengolahan maka data tersebut telah dapat diolah atau diproses menggunakan proses klusterisasi (DQLab, 2021). Berikut adalah variabel atau atribut yang digunakan dalam proses optimasi.

Tabel 3.1 Data Attribute

Variabel	Tipe Data
Curah Hujan	Numerikal
Kehilangan Produksi	Numerikal
Bencana Alam	Numerikal

Selanjutnya berdasarkan penjelasan pada table 3.1 kemudian dilakukan perekaman data berdasarkan atribut tersebut. Perekaman data dilakukan pada masing-masing kecamatan dalam wilayah Kabupaten Jember.

Tabel 3.2 Data per Kecamatan dalam wilayah Kabupaten Jember

No	Kecamatan	Curah Hujan	Kehilangan Produksi	Bencana Alam
1	Ajung	0	36,43	1
2	Ambulu	10	141,7	5
3	Arjasa	11	29,17	1
4	Balung	12	8,62	1
5	Bangsalsari	13	39,03	3
6	Gumukmas	9	26,48	0
7	Jelbuk	14	19,63	2
8	Jenggawah	11	24,97	0
9	Jombang	7	18,45	2
10	Kalisat	12	26,45	2
11	Kaliwates	11	8,04	18
12	Kencong	10	17,38	1
13	Ledokombo	16	64,81	5
14	Mayang	13	64,78	8
15	Mumbulsari	0	11,32	2
16	Pakusari	9	3,52	0
17	Panti	11	25,46	3
18	Patrang	10	3,71	11
19	Puger	11	25,22	2

No	Kecamatan	Curah Hujan	Kehilangan Produksi	Bencana Alam
20	Rambipuji	10	42,49	5
21	Sembo	8	152,94	0
22	Silo	12	84,69	6
23	Sukorambi	10	17,79	4
24	Sukowono	14	54,66	2
25	Sumber Baru	11	54,89	2
26	Sumber Jambe	12	72,16	2
27	Sumbersari	10	16,47	10
28	Tanggul	14	148,34	1
29	Tempurejo	12	476,96	2
30	Umbulsari	8	0,23	0
31	Wuluhan	11	97,09	4

Data yang telah diperoleh selanjutnya dilakukan normalisasi untuk mendapatkan keseimbangan data dari masing-masing atribut. Proses normalisasi dilakukan dengan menentukan nilai Z-score dari masing-masing kecamatan per atribut.

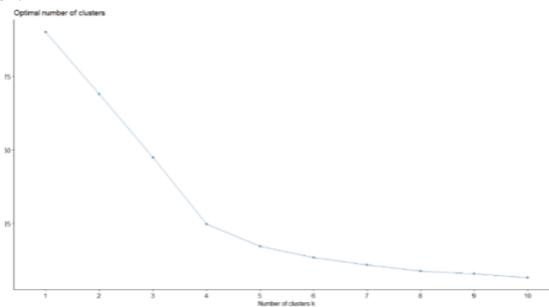
Tabel 3.3 Normalisasi Data

No	Kecamatan	Curah Hujan	Kehilangan Produksi	Bencana Alam
1	Ajung	-3,075	-0,250	-0,613
2	Ambulu	-0,042	0,942	0,414
3	Arjasa	0,188	-0,332	-0,613
4	Balung	0,575	-0,565	-0,613
5	Bangsalsari	0,837	-0,221	-0,099
6	Gumukmas	-0,302	-0,363	-0,869
7	Jelbuk	1,078	-0,440	-0,356
8	Jenggawah	0,175	-0,380	-0,869
9	Jombang	-0,966	-0,454	-0,356
10	Kalisat	0,484	-0,363	-0,356
11	Kaliwates	0,188	-0,572	3,751
12	Kencong	-0,081	-0,466	-0,613
13	Ledokombo	1,708	0,071	0,414
14	Mayang	0,645	0,071	1,184
15	Mumbulsari	-3,075	-0,535	-0,356
16	Pakusari	-0,300	-0,623	-0,869
17	Panti	0,056	-0,374	-0,099
18	Patrang	-0,035	-0,621	1,954
19	Puger	0,048	-0,377	-0,356
20	Rambipuji	-0,074	-0,182	0,414
21	Sembo	-0,721	1,070	-0,869
22	Silo	0,348	0,297	0,671
23	Sukorambi	-0,109	-0,461	0,157
24	Sukowono	1,127	-0,044	-0,356
25	Sumber Baru	0,173	-0,041	-0,356

No	Kecamatan	Curah Hujan	Kehilangan Produksi	Bencana Alam
26	Sumber Jambe	0,484	0,155	-0,356
27	Sumbersari	-0,257	-0,476	1,697
28	Tanggul	1,037	1,018	-0,613
29	Tempurejo	0,484	4,740	-0,356
30	Umbulsari	-0,715	-0,660	-0,869
31	Wuluhan	0,117	0,437	0,157

3.3 Menentukan nilai K Optimal

Dalam penentuan nilai atau jumlah k terbaik, dalam penelitian ini nilai Elbow yang akan disajikan dalam bentuk grafik menggunakan nilai SSE. Hal tersebut dilakukan bertujuan untuk mendapatkan nilai k optimum dari hasil perbandingan dengan nilai k lainnya. Dalam proses ini berdasarkan nilai perbandingan dari nilai k=1 hingga k=10 menunjukkan bahwa kondisi elbow atau k optimal ditemukan pada nilai k=4. Hal tersebut menunjukkan bahwa kondisi optimal dari proses klusterisasi ditemukan pada nilai k=4 atau data dapat dikelompokkan pada 4 cluster.



Gambar 3.1 Hasil Elbow

3.4 Menentukan Centroid

Dalam proses penentuan nilai k optimal berdasarkan nilai SSE (Sum Square Error) dapat dilakukan saat data-data yang digunakan telah berkelompok. Pengelompokan data dapat dilakukan dengan cara menghitung jarak antar data atau kemiripan dari masing-masing data dengan menggunakan nilai Euclidean Distance. Oleh karena itu untuk estimasi awal dari proses pengelompokan atau klusterisasi diperlukan nilai estimasi awal dari centroid (titik pusat) cluster.

Tabel 4.1 Estimasi Awal Nilai Centroid (k=4)

Centroid			
Cluster	x	y	z
1	-0,071	-0,201	-0,452
2	-0,175	-0,359	0,350
3	0,080	-0,087	0,189
4	0,189	0,739	-0,099

Selanjutnya berdasarkan nilai centroid awal ditentukan keanggotaan dari masing-masing cluster. Setelah data berkelompok maka dilanjutkan dengan proses pengujian nilai dan perbandingan nilai SSE dari masing-masing cluster untuk mengetahui apakah cluster yang terbentuk masih terdapat kondisi outliers atau tidak. Jika kondisi outliers masih ditemukan maka nilai centroid dirubah hingga kondisi outliers tidak ditemukan.

Tabel 4.2 Nilai Centroid Final (k=4)

Centroid			
Cluster	x	y	z
1	-3,075	-0,392	-0,484
2	0,135	-0,399	2,147
3	0,214	-0,098	-0,303
4	0,484	4,740	-0,356

4. Kesimpulan

Berdasarkan hasil dari proses klusterisasi kecamatan-kecamatan di Kabupaten Jember menghasilkan 4 kelompok yang memiliki nilai jumlah keanggotaan yang berbeda. Selanjutnya pada proses tersebut perulangan yang dilakukan dengan menggunakan nilai $k = 1$ hingga $k = 10$. Berdasarkan hasil uji nilai SSE dan grafik menunjukkan nilai k optimum ditemukan pada iterasi $k=4$ dengan nilai SSE sebesar 24,809.

Daftar Pustaka

- [1] Putu, N., E. Merliana, dan A. J. Santoso. 2016. Analisa penentuan jumlah cluster terbaik pada metode k-means. PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK (SENDI_U). 978–979.
- [2] Ananda, R. dan A. Z. Yamani. 2021. Penentuan centroid awal k-means pada proses clustering data evaluasi. *JURNAL RESTI (Rekayasa Sistem Dan Teknologi Informasi)*. 1(10):544–550.
- [3] BBNP. 2007. Definisi Bencana. <https://bnpb.go.id/Definisi-Bencana>
- [4] BBNP. 2012. Potensi ancaman bencana
- [5] Faisal, M., E. M. Zamzami, dan Sutarman. 2020. Comparative analysis of inter-centroid k-means performance using euclidean distance, canberra distance and manhattan distance. *Journal of Physics: Conference Series*. 1566(1)
- [6] Kabupaten Jember dalam Angka. 2020. Dalam angka dalam angka. *Kabupaten Jember Dalam Angka 2021*. 1–68.
- [7] Pangan, B. K. dan K. Pertanian. 2020. PANDUAN penyusunan peta ketahanan dan kerentanan pangan (*food security and vulnerability atlas/fsva*) provinsi 2020
- [8] Tan, P. N., Steinbach, M., dan Kumar, V., "Introduction to Data Mining", 1st Edition, 2004.
- [9] Yana, M. S., L. Setiawan, E. M. Ulfa, dan A. Rusyana. 2018. Penerapan metode k-means dalam pengelompokan wilayah menurut intensitas kejadian bencana alam di indonesia tahun 2013-2018. *Journal of Data Analysis*. 1(2):93–102.
- [10] DQLab. 2021. Pentingnya Preprocessing Dalam Pengolahan Data Statistik. <https://www.dqlab.id/pentingnya-preprocessing-dalam-pengolahan-data-statistik>
- [11] Oktari, Y. 2018. Pengaruh bencana alam terhadap ketahanan pangan di kabupaten ogan komering ilir. (May)
- [12] Hapsari, N. I. dan I. Rudiarto. 2017. Faktor-faktor yang mempengaruhi kerawanan dan ketahanan pangan dan implikasi kebijakannya di kabupaten rembang. *Jurnal Wilayah Dan Lingkungan*. 5(2):125.
- [13] Tirivangasi, H. M. 2018. Regional disaster risk management strategies for food security: probing southern african development community channels for influencing national policy. *Jamba: Journal of Disaster Risk Studies*. 10(1):1–7.
- [14] Pangan, B. K. dan K. Pertanian. 2020. PANDUAN penyusunan peta ketahanan dan kerentanan pangan (*food security and vulnerability atlas/fsva*) provinsi 2020
- [15] Mulyana Nasution, F. 2019. Penerapan Metode K-Means Clustering Untuk Mengelompokkan Ketahanan Pangan Kabupaten/Kota Di Provinsi Sumatera Utara. Analisis Kesadahan Total Dan Alkalinitas Pada Air Bersih Sumur Bor Dengan Metode Titrimetri Di PT Sucofindo Daerah Provinsi Sumatera Utara.