

## Terakreditasi SINTA Peringkat 4

Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 28/E/KPT/2019 masa berlaku mulai Vol.3 No. 1 tahun 2018 s.d Vol. 7 No. 1 tahun 2022

Terbit online pada laman web jurnal:  
<http://publishing-widyagama.ac.id/ejournal-v2/index.php/jointecs>



Vol. 6 No. 1 (2021) 13 - 20

# JOINTECS

## (Journal of Information Technology and Computer Science)

e-ISSN:2541-6448

p-ISSN:2541-3619

### Klasterisasi Menggunakan *Agglomerative Hierarchical Clustering* Untuk Memodelkan Wilayah Banjir

Ridzki Okta Pratiko<sup>1</sup>, Natalia Damastuti<sup>2</sup>

<sup>1</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Narotama

<sup>2</sup>Program Studi Sistem Komputer, Fakultas Ilmu Komputer, Universitas Narotama

<sup>1</sup>ridzki.okta@mhs.fasilkom.narotama.ac.id, <sup>2</sup>natalia.damastuti@narotama.ac.id

#### Abstract

Every year during the rainy season the problem of flooding in the province of East Java is a frequent disaster. Based on the records of the Disaster Management Agency (BNPB) from 2014 to 2015 there were 574 flood disasters in the province of East Java. Many factors cause floods, including the slow information obtained, so we need a modeling of potential flood areas in East Java using a more accurate and efficient method using an Agglomerative Hierarchical Clustering (AHC) method. This method will be used to group performance tests using the silhouette score method and cophenetic correlation coefficient. The results of this study are visualized in the form of GIS. Based on the optimal cluster results with the elbow method, the province of East Java is divided into 3 clusters of areas affected by flood potential i.e. low, medium, and high characteristics. Cluster performance results using cophenetic correlation coefficient indicate that the average linkage method provides a better cluster solution compared to other AHC methods which is 0.92.

Keywords: floods; clustering; agglomerative; elbow method; cophenetic correlation coefficient.

#### Abstrak

Setiap tahun selama musim hujan masalah banjir di provinsi Jawa Timur adalah bencana yang sering terjadi. Berdasarkan catatan Badan Penanggulangan Bencana (BNPB) dari 2014 hingga 2015 ada 574 bencana banjir di provinsi Jawa Timur. Banyak faktor yang menyebabkan terjadinya bencana banjir diantaranya adalah lambatnya informasi yang didapat sehingga diperlukan suatu pemodelan wilayah potensi banjir di Jawa Timur dengan menggunakan metode yang lebih akurat dan efisien dengan menggunakan suatu metode *Agglomerative Hierarchical Clustering* (AHC). Metode ini akan digunakan untuk melakukan pengelompokan dengan uji performansi menggunakan metode *cophenetic correlation coefficient*. Hasil dari penelitian ini divisualisasikan kedalam bentuk SIG. Berdasarkan hasil uji cluster optimal dengan *elbow method*, provinsi Jawa Timur terbagi menjadi 3 kelompok *cluster* daerah terdampak potensi banjir yaitu karakteristik rendah, sedang, tinggi. Hasil uji performa cluster menggunakan *cophenetic correlation coefficient* menunjukkan bahwa metode average linkage memberikan solusi cluster yang lebih baik dibandingkan dengan metode AHC lainnya yakni sebesar 0,92.

Kata kunci: banjir; klasterisasi; *agglomerative*; metode *elbow*; *cophenetic correlation coefficient*.

© 2021 Jurnal JOINTECS

#### 1. Pendahuluan

Pulau Jawa merupakan wilayah dengan pertumbuhan ekonomi yang pesat sehingga terdapat hambatan dalam hal transportasi apabila infrastruktur kurang memadai. Setiap tahun saat musim hujan permasalahan banjir di provinsi Jawa Timur merupakan bencana yang sering

terjadi. Banjir adalah suatu kejadian dimana air di dalam saluran meningkat dan melampaui kapasitas daya tampungnya [1]. Berdasarkan data BNPB telah tercatat pada tahun 2014 sampai 2018 bencana banjir di provinsi Jawa Timur telah terjadi sebanyak 574 kejadian. Faktor penyebab terjadinya banjir menurut [2] yaitu: banjir yang disebabkan oleh hujan lebat, kurangnya daerah

Diterima Redaksi : 05-07-2020 | Selesai Revisi : 27-09-2020 | Diterbitkan Online : 31-01-2021

resapan air seperti kurangnya pohon, perilaku masyarakat yang sering membuang sampah ke sungai menyebabkan kenaikan permukaan air sungai, dan fasilitas pengendalian banjir yang tak mampu menahan air hujan. Upaya pencegahan bencana banjir pada daerah yang berpotensi terdampak perlu dilakukan untuk memperkecil dampak yang disebabkan salah satunya dengan strategi pemetaan dampak bencana banjir pada provinsi Jawa Timur [3].

Berdasarkan uraian di atas maka perlu adanya pengelompokan wilayah kabupaten/kota di Jawa Timur berdasarkan kesamaan dampak kejadian bencana banjir yang ditimbulkan. Guna dapat memetakan dampak bencana banjir pada wilayah jawa timur peneliti menggunakan salah satu teknik data mining yaitu *clustering* yaitu untuk melakukan pengelompokan data menjadi objek yang serupa sesuai dengan ciri maupun karakteristiknya[4]. Penelitian ini menggunakan Algoritma *agglomerative hierarchical clustering* (AHC) yang merupakan salah satu metode *clustering* yang mengelompokkan objek data ke dalam sebuah hirarki kelompok data menggunakan strategi *bottom-up* [5]. Hal ini juga dimaksudkan untuk mengetahui efisiensi algoritma AHC. Berdasarkan pada faktor penyebab banjir maka dalam penelitian ini menggunakan parameter input dari data curah hujan yang berasal dari Badan Meteorologi Klimatologi dan Geofisika (BMKG), data jumlah kejadian bencana banjir selama 5 tahun yang dicatat oleh Badan Nasional Penanggulangan Bencana (BNPB), data kepadatan penduduk dari Badan Pusat Statistika (BPS), korban terdampak mengungsi dari BNPB, dan data rumah terendam dari BNPB [6].

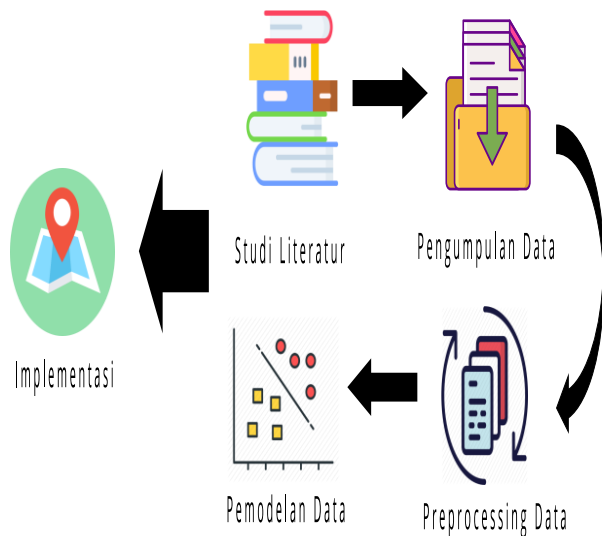
Terdapat beberapa penelitian yang pernah dilakukan tentang pengelompokan wilayah terdampak potensi banjir. Pada [7] menganalisis daerah rawan banjir di bangladesh berdasarkan catatan kejadian banjir dan dengan membandingkan tiga metode berbeda yaitu *k-means*, *expectation maximation* dan *agglomerative clustering*, hasil menunjukkan metode *k-means* menghasilkan akurasi lebih baik dengan ketepatan model 17,5% yang menandai pengaruh signifikan sungai india dan curah hujan terhadap banjir di bangladesh. Penulis [8] menganalisis intensitas peristiwa banjir di sungai wujiang, cina selatan pada lebih 53 tahun dengan 5 indikator puncak debit, tingkat puncak, volume maksimum 24 jam, volume maksimum 72 jam dan total volume banjir, menggunakan algoritma *fuzzy c-means* menunjukkan bahwa tingkat banjir tinggi terjadi sejak pada tahun 1990-an di daerah aliran sungai. Pada penelitian [9] melakukan analisis *clustering* dampak bencana alam menggunakan algoritma *k-means*, hasil penelitian yang dilakukan menunjukkan bahwa kondisi cuaca bukanlah penyebab utama terjadinya bencana alam, namun kondisi geografis merupakan pemicu utama masalah bencana alam.

Pada penelitian ini akan dibagi 3 kelompok klaster daerah terdampak banjir dengan tujuan untuk

mengetahui karakteristik dampak banjir yang tinggi, sedang, dan rendah kemudian hasil klaster tersebut digunakan sebagai acuan menandai lokasi kabupaten/kota jawa timur untuk ditampilkan ke dalam sistem informasi geografis. Adapun tujuan penelitian diharapkan dapat membantu pemerintah memetakan dampak daerah bencana banjir pada wilayah kabupaten/kota jawa timur berdasarkan kesamaan kejadian bencana banjir sebelumnya. Tujuan yang lain untuk mengetahui ketepatan *cluster* optimal yang dihasilkan algoritma AHC dengan analisis *cophenetic correlation coefficient*.

## 2. Metode Penelitian

Pada penelitian ini terdapat tahapan dan alur metode penelitian yang akan dilakukan untuk menyelesaikan masalah dalam penelitian ini. Kerangka alur penelitian ini disajikan pada Gambar 1. Berdasarkan alur penelitian dilakukan tahapan proses antara lain studi literatur, pengumpulan data, preprocessing data, pemodelan data, dan implementasi.



Gambar 1. Metodologi penelitian

### 2.1. Studi Literatur

Seperti yang ditunjukkan pada Gambar 1 penelitian dimulai dengan mencari referensi teori yang relevan pada kasus pengelompokan dampak potensi bencana banjir dari jurnal, artikel, buku. Tujuan dari referensi tersebut akan digunakan sebagai dasar teori untuk menyelesaikan permasalahan pada penelitian ini. Referensi tersebut berisikan tentang banjir secara umum, pemrosesan sebuah data mentah, algoritma *agglomerative hierarchical clustering*, dan implementasi sistem informasi geografis dari hasil pemodelan dengan *agglomerative hierarchical clustering*.

### 2.2. Pengumpulan Data

Kedua mengumpulkan data, pengumpulan data bertujuan untuk mengumpulkan informasi yang akan digunakan untuk menyelesaikan masalah. Lokasi penelitian yang dipilih adalah seluruh kabupaten/kota

provinsi Jawa Timur. Data yang dikumpulkan merupakan data curah hujan yang berasal dari BMKG, data kepadatan penduduk yang diperoleh dari BPS, dan data jumlah kejadian banjir, jumlah korban terdampak dan mengungsi, rumah terendam yang dicatat oleh BNPB. Data yang dikumpulkan selama 5 tahun terakhir dari 2014 sampai 2018 dalam bentuk format excel.

### 2.3. Preprocessing Data

Tahap yang ketiga selanjutnya data yang telah diperoleh dilakukan *preprocessing* pada data. Data *Preprocessing* dilakukan karena masih terdapat isian data yang belum terisi, dan juga masih terdapat kemungkinan sebaran data masih belum seimbang. Dalam penelitian ini terdapat 3 tahapan *preprocessing* yang terdiri dari data *cleaning*, data *integration*, dan data *transformation*.

Data *cleaning* dilakukan pada data curah hujan dimana diantaranya masih terdapat data yang diisi dengan nilai 8888 yang menandakan data tidak terukur dan isi nilai data 9999 yang berarti tidak dilakukan pencatatan. Sehingga untuk menghasilkan analisis yang lebih akurat perlu untuk menghilangkan isian data yang diisi dengan nilai 8888 dan 9999. Sedangkan data yang kosong akan diisi dengan nilai rata-rata dari seluruh data.

Setelah proses data *cleaning* langkah selanjutnya adalah data *integration* yang bertujuan untuk menyatukan dua atau lebih atribut yang berasal dari berbagai sumber data agar lebih efisien. Pada penelitian ini tiga sumber data, yaitu data curah hujan, data kepadatan penduduk, data kejadian banjir, data korban mengungsi, dan data rumah terendam akan digabung menjadi satu dataset. Dalam menggabungkan data, peneliti mengelompokkan data berdasarkan tahunan.

Selanjutnya melakukan data *transformation*, hal ini dilakukan agar nilai pada atribut sebaran data menjadi seimbang sehingga data sesuai untuk proses data mining. Salah satu teknik data transformasi yang digunakan dalam penelitian ini adalah melakukan normalisasi data dengan *Z-score*. Secara matematis perhitungan *Z-score* dapat menggunakan rumus pada rumus 1 [10]:

$$z = \frac{x_i - \mu}{\sigma} \quad (1)$$

Keterangan:

$z$  = nilai *Z-score*

$x$  = nilai yang diamati

$\mu$  = rata-rata

$\sigma$  = standar deviasi

### 2.4. Pemodelan Data

Data mentah yang telah diolah melalui tahap *preprocessing* selanjutnya akan dilakukan analisis menggunakan metode *clustering*. Pada penelitian ini metode *clustering* digunakan untuk mengelompokkan data banjir dengan kesamaan karakteristik dampak terjadi banjir. Algoritma *clustering* yang digunakan dalam penelitian ini adalah *Agglomerative Hierarchical Clustering*. Pemodelan *clustering* penelitian ini terdiri

dari 3 tahapan yaitu menentukan jumlah *cluster*, proses *cluster*, ukur kualitas *cluster*.

#### 2.4.1. Menentukan Jumlah Cluster

Sebelum melakukan proses *clustering* terlebih dahulu menentukan berapa banyak jumlah *cluster* optimal untuk dibentuk [4]. Pada penelitian ini *elbow method* digunakan dalam menentukan jumlah *cluster* optimal. Gambar 2 merupakan algoritma dalam menentukan nilai  $k$  menggunakan *elbow method*. *Elbow method* didasarkan pada pengamatan grafik yang dihasilkan pada setiap peningkatan jumlah *cluster* yang dibentuk dengan menggunakan perhitungan SSE (*Sum of Square of Error*) pada setiap skema *cluster* hingga membentuk siku pada suatu titik [11]. Berdasarkan nilai SSE grafik *elbow method* menggunakan rumus 2:

$$SSE = \sum_{i=1}^n (X_i - \bar{X}) \quad (2)$$

Keterangan:

SSE = *Sum of Square of Error*

$X_i$  = Titik data  $x$  pada indeks ke- $i$

$\bar{X}$  = Rata-rata semua titik data pada *cluster*.

Dua titik data yang memiliki nilai jarak yang dekat maka nilai perhitungan menggunakan rumus 2 akan mendapatkan nilai SSE yang kecil dan sebaliknya apabila dua titik data memiliki jarak yang jauh maka akan mendapatkan nilai SSE yang besar. Optimalnya SSE yang dihasilkan memiliki nilai kecil dan jumlah *cluster* yang sedikit agar akurasi yang dihasilkan baik sehingga dapat berguna untuk menarik kesimpulan pada setiap *cluster*. Apabila nilai SSE yang dihasilkan oleh suatu *cluster* cukup besar maka *cluster* yang dihasilkan memiliki perbedaan.

#### 2.4.2. Clustering Menggunakan AHC

Pada tahap ini akan dilakukan proses *clustering* dengan menggunakan metode AHC. Dalam AHC terdapat 3 metode *cluster* hirarki yang dapat digunakan yaitu metode *single linkage*, *complete linkage*, dan *average linkage*. *Single linkage* adalah proses pengklasteran yang didasarkan pada jarak antara anggota-anggota yang terdekat. *Complete linkage* adalah proses pengklasteran yang didasarkan pada jarak antara anggota-anggota yang terjauh. *Average linkage* adalah proses pengklasteran berdasarkan jarak rata-rata antar objeknya dari semua pasangan titik data dalam dua klaster [12].

Pada penelitian ini akan digunakan metode *average linkage* untuk mengelompokkan dampak bencana banjir. Dalam pengklasteran dengan metode *average linkage* pengelompokkan klaster dimulai dari mengubah data ke dalam sebuah matriks jarak  $D = \{dik\}$  untuk memperoleh objek paling dekat misalnya U dan V. Terdapat beberapa cara untuk mengukur jarak antar objek, metode yang paling umum biasa digunakan diantaranya adalah menggunakan perhitungan jarak *Euclidean*, jarak *minkowski/manhattan*, jarak *mahalanobis*, dan jarak *Canberra*. Pada penelitian ini akan menggunakan salah satu cara dengan ukuran jarak *Euclidean*. Jarak

*Euclidean* adalah salah satu metode pemetaan objek yang mempunyai kemiripan dalam analisis *cluster* dengan melakukan pengukuran jarak antara dua titik [13]. Secara matematis jarak *Euclidean* dapat diformulasikan dalam rumus 3 [13].

$$D(x, y) = \sqrt{\sum_{k=1}^n (X_k - Y_k)^2} \quad (3)$$

Keterangan:  
 $D(x,y)$  = Jarak *Euclidean* objek data ke- $i$  dan objek data ke- $j$   
 $n$  = banyak parameter yang digunakan  
 $X_k$  = objek data ke- $k$   
 $Y_k$  = objek data ke- $k$

Hasil perhitungan *euclidean distance* menggunakan rumus 3 akan diperoleh matriks jarak  $D=\{d_{ik}\}$  yang kemudian menggabungkan objek objek yang dekat atau mirip untuk mendapatkan kluster UV. Misalnya U dan V untuk mendapatkan kluster (UV). Kemudian jarak-jarak antara (UV) dan *cluster* W yang lain ditentukan dengan rumus 4.

$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_{(uv)} N_w} \quad (4)$$

Keterangan:  
 $d_{(uv)w}$  = jarak *cluster* dengan titik data  $w$   
 $d_{ik}$  = jarak antara titik data  $i$  dalam *cluster* (UV) dan objek  $k$  dalam *cluster* W  
 $N_{(uv)}$  = banyak titik data dalam *cluster* (UV)  
 $N_w$  = banyak titik data dalam *cluster* W

#### 2.4.3. Ukur Kualitas Cluster

Tahap selanjutnya akan dilakukan evaluasi yang merupakan tahap pengukuran kualitas *cluster* yang dihasilkan dengan menggunakan uji performa berdasarkan metode *cophenetic correlation coefficient*. *Cophenetic correlation coefficient* adalah sebuah metode yang umum digunakan untuk mengukur *cluster* yang dihasilkan pada algoritma *hierarchical clustering*. *Cophenetic correlation coefficient* memiliki rentang nilai dari 0 sampai 1, apabila nilai yang dihasilkan semakin besar atau mendekati 1 maka metode *linkage* yang digunakan sudah tepat untuk digunakan dalam *clustering* [14]. Perhitungan *cophenetic correlation coefficient* pada rumus 5

$$c = \frac{\sum_{i < j} (x(i,j) - \bar{x})(t(i,j) - \bar{t})}{\sqrt{[\sum_{i < j} (x(i,j) - \bar{x})^2][\sum_{i < j} (t(i,j) - \bar{t})^2]}} \quad (5)$$

Keterangan:  
 $c$  = nilai *cophenetic correlation coefficient*  
 $x(i,j)$  = *Euclidean distance* antara *cluster* ke  $i$  dan  $j$   
 $t(i,j)$  = *dendogram distance* dari *cluster*  $i$  dan  $j$   
 $\bar{x}$  = rata-rata dari  $x(i,j)$

#### 2.5. Implementasi

Pada tahap ini dilakukan pembuatan sistem dari hasil analisis *cluster* yang sudah dilakukan dalam bentuk aplikasi berbasis *web*. Alat yang digunakan dalam membangun aplikasi berbasis *web* menggunakan bahasa

pemrograman *python* dengan *framework flask*. Pemrograman *python* adalah pemrograman komputer open source untuk keperluan umum yang dioptimalkan untuk kualitas, produktivitas, portabilitas, dan integrasi, juga memungkinkan untuk membantu penelitian dalam melakukan analisis data [15]. *Framework flask* adalah *web framework* untuk bahasa pemrograman *python* yang memudahkan pengguna *python* dalam mengembangkan aplikasi berbasis *web*.

### 3. Hasil dan Pembahasan

Pada bagian ini akan dijabarkan dalam implementasi hasil dari perancangan yang telah disusun pada metodologi penelitian. Pada penelitian ini menggunakan data yang telah tercatat dan dikumpulkan selama 5 tahun terakhir di provinsi Jawa Timur dengan total data yang diperoleh sebanyak 950 data dengan menggunakan atribut curah hujan, jumlah kepadatan penduduk, riwayat banjir, jumlah korban terdampak pengungsi, jumlah rumah terendam. Proses analisis *clustering* dalam penelitian ini menggunakan pemrograman *python*. Berikut dijelaskan hasil dari *preprocessing*, hasil dari pembentukan model dengan AHC, hingga hasil visualisasi hasil analisis *cluster* dalam bentuk sistem informasi geografis.

#### 3.1. Preprocessing

Tahap awal yang dilakukan adalah melakukan pembersihan pada data curah hujan. Setelah menyelesaikan proses data *cleansing* kemudian melakukan proses penggabungan data atau data *integration* dimana data yang telah dikumpulkan dari berbagai sumber digabung menjadi satu dataset hasilnya 950 *records* data dikumpulkan yang terdiri dari atribut curah hujan, kepadatan penduduk, jumlah riwayat banjir, jumlah korban terdampak, dan jumlah rumah terendam. Kemudian data ditransformasi menggunakan Z-score dengan menggunakan rumus 1 agar setiap nilai atribut menjadi seimbang. Berikut hasil akhir dari tahapan *preprocessing* data dapat dilihat pada Tabel 1.

Tabel 1. Hasil *Preprocessing*

Curah Hujan	Kepadatan Penduduk	Riwayat banjir	Korban mengungsi	Rumah terendam
-0.16327	-0.55019	-0.34437	-0.35805	-0.17326
-0.39822	-0.72711	-0.6651	-0.03905	-0.39665
-0.41003	-0.55964	-0.98584	-0.36623	-0.39665
-0.40358	-0.59291	1.580044	-0.09697	-0.04309
-0.3947	-0.64772	-0.98584	-0.36623	-0.39665
-0.39068	-0.41221	-0.34437	-0.36623	0.347032
-0.38369	-0.55699	0.938573	0.208553	0.827519
-0.27659	-0.42846	-0.6651	0.497057	-0.39665
-0.39565	-0.42392	-0.98584	-0.36623	-0.39665
0.349127	2.512305	-0.98584	-0.36623	-0.39665

#### 3.2. Pemodelan Data

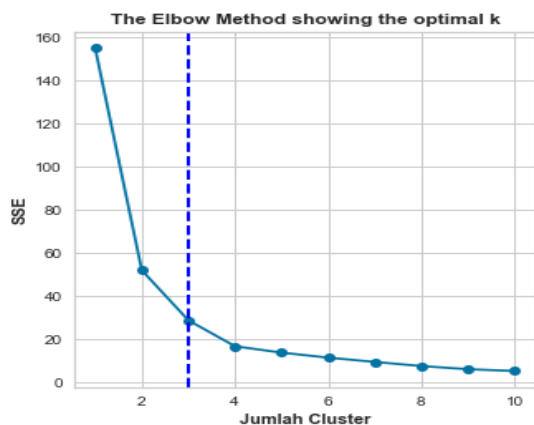
Data mentah yang telah diolah dan melalui tahap *preprocessing* kemudian akan dianalisis menggunakan metode *clustering*. Sebelum melakukan analisis

*clustering* terlebih dahulu untuk menentukan jumlah *cluster* optimal yang akan diterapkan menggunakan *elbow method* dengan melakukan iterasi sebanyak 10 kali K=0 sampai dengan K=9 menggunakan perhitungan SSE. Hasil iterasi SSE dari setiap iterasi menggunakan persamaan rumus 2 dapat dilihat pada Tabel 2.

Tabel 2. Hasil SSE dengan K=10

Iterasi ke-i	SSE
0	155.037931
1	51.967615
2	28.709338
3	16.684522
4	13.672755
5	11.190530
6	8.960330
7	7.498220
8	6.046961
9	5.041667

Berdasarkan pada Tabel 2, hasil SSE dengan jumlah K=0 memiliki nilai 155.03 selanjutnya dengan K=1 memiliki SSE senilai 51.96 lalu ketika dibagi dengan K=2 maka nilai SSE sebesar 28.71 sementara bila dikelompokkan dalam K=3 nilai SSE akan turun senilai 16.684 dan seterusnya. *Cluster* optimal dari perhitungan SSE sebesar 16.684 pada iterasi ke-3, dengan alasan karena nilai SSE yang dihasilkan relative kecil dan jumlah *cluster* yang tidak terlalu banyak. Sehingga dalam penelitian ini peneliti membentuk *cluster* menjadi 3 kelompok dengan tujuan untuk melihat wilayah mana yang memiliki dampak potensi banjir tinggi, sedang, rendah. Perhitungan hasil SSE dari setiap iterasi dapat dilihat lebih jelas menggunakan *grafik line* pada Gambar 2.



Gambar 2. Grafik Elbow Method

Pada Gambar 2 dapat dilihat dimana sumbu x adalah jumlah *cluster* yang ada dan sementara sumbu y adalah SSE pada setiap iterasi *cluster*. Apabila dilihat pada grafik Gambar 2 terdapat suatu titik yang memiliki sudut yang ditandai pada garis potong vertikal yang merupakan titik optimal untuk dipilih yaitu jumlah *cluster* 3. Atas dasar hasil grafik dari perhitungan SSE maka untuk penyelesaian dalam penelitian jumlah

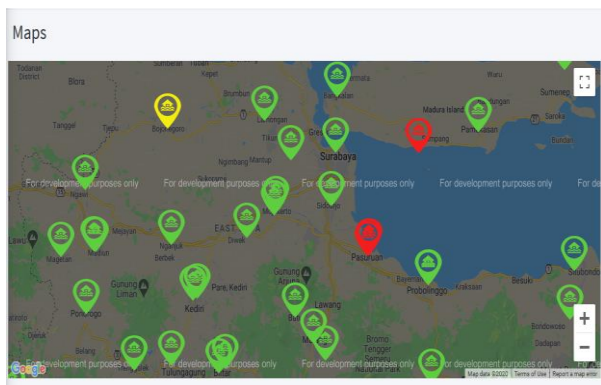
*cluster* yang ideal adalah K=3 untuk menentukan karakteristik dari data penelitian yang digunakan pada penelitian ini.

Selanjutnya melakukan analisis kluster menggunakan algoritma *agglomerative hierarchical clustering* dengan salah satu metode *linkage* yaitu *average linkage*. Pada penelitian ini, dilakukan proses *clustering* mulai dari tahun 2014 sampai 2018. Dalam menentukan asumsi kategori *cluster* aman, cukup rawan, dan rawan dapat dilihat dari hasil perhitungan rata-rata variabel yang tertinggi dan terendah secara keseluruhan dari setiap atribut curah hujan (CH), kepadatan penduduk (KP), riwayat banjir (RB), korban mengungsi (KM), rumah terendam (RT). Hasil analisis perhitungan AHC metode *average linkage* disajikan dalam Tabel 3.

Tabel 3. Profilisasi Hasil *Cluster*

Kluster	N	Rata-rata				
		CH	KP	RB	KM	RT
C1	1	93.23	172	25	2370	2600
C2	34	3055.19	2188.4	2.8	2858.11	622.88
C3	3	1915.73	2677.5	6.625	67964	11854

Hasil analisis *cluster* pada Tabel 3 memperlihatkan kelompok kluster C1 hanya beranggotakan 1 kabupaten saja yaitu Kabupaten Bojonegoro, dimana pada daerah tersebut tergolong daerah kategori sedang atau cukup rawan terjadi banjir. Wilayah pada kluster C1 berada di daerah yang sebagian besar memiliki keadaan tanah berbukit sedangkan permukaan tanah rata-rata relatif datar dan di lalui sungai bengawan solo. Namun berdasarkan Tabel 3 faktor yang mempengaruhi hasil kelompok kluster C1 termasuk sedang atau cukup rawan adalah jumlah riwayat kejadian banjir sebanyak 25. Sementara itu di daerah pada kelompok kluster C2 mayoritas merupakan daerah tergolong aman. Kluster C2 sebagian besar banyak dikelompokkan pada daerah dataran rendah dan dekat pegunungan seperti daerah yang berada disekitar pegunungan pada Kabupaten/Kota Bondowoso, Jember, Lumajang, Madiun, Magetan, Malang, Nganjuk, Ngawi, Pacitan, Probolinggo, Situbondo, Trenggalek, Tulungagung, Kota Batu, Kabupaten Blitar, Kota Malang, Kota Probolinggo. Sedangkan dataran rendah ada pada Kota Surabaya, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Kediri, Kabupaten Kediri, Lamongan, Tuban, Sumenep, Sidoarjo, Ponorogo, Pamekasan, Bangkalan, Banyuwangi, Kota Blitar, Gresik, Jombang. Faktor yang mempengaruhi hasil kelompok C2 termasuk aman adalah nilai jumlah riwayat banjir dan rumah yang terendam yang rendah. Kluster C3 merupakan kelompok yang tergolong rawan. Kluster C3 banyak di kelompokkan di daerah yang dekat dengan pesisir pantai seperti pada Kabupaten Pasuruan, Kota Pasuruan, dan Kabupaten Sampang. Dengan jumlah korban yang mengungsi dan rumah terendam cukup banyak. Hasil dari analisis ini kemudian ditampilkan dalam sistem informasi geografis pada Gambar 3.



Gambar 3. Hasil Pemetaan GIS

Pada Gambar 3 disajikan tampilan sistem informasi geografis dapat dilihat terdapat 3 tanda berupa warna yang berbeda yaitu warna kuning yang berarti kluster C1, warna hijau kluster C2, warna merah kluster C3. Kondisi pewarnaan ini didasarkan pada hasil analisis dari perhitungan rata-rata variabel pada setiap kluster. Sehingga pada Gambar 3 diketahui bahwa Kabupaten Bojonegoro daerah dengan kluster C1 dengan kondisi cukup rawan. kluster C3 pada daerah Kabupaten Pasuruan, Kota Pasuruan, dan Kota Sampang dengan kondisi rawan. Daerah lainnya ada pada kluster C2 termasuk dalam kondisi aman.

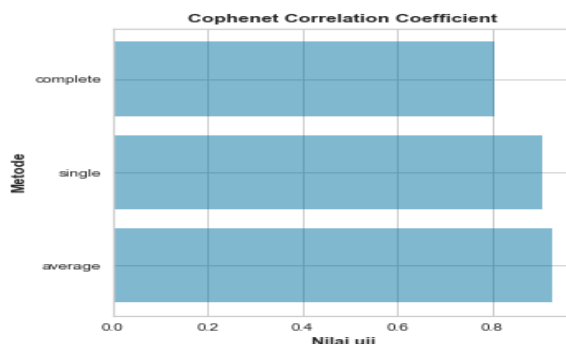
Proses selanjutnya yaitu pengujian performa *cluster* pada metode *linkage* yang digunakan. dalam penelitian ini uji performa menggunakan metode *coefficient cophenetic correlation*, dengan menggunakan persamaan pada rumus 5. Tujuan dari pengujian ini untuk mengetahui ketepatan metode hierarki *clustering* yang digunakan yang terdiri dari metode *single linkage*, *complete linkage*, *average linkage*. Hasil uji performa dengan *cophenetic correlation coefficient* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Uji Performa

Metode linkage	Nilai uji
Complete	0.805725
Single	0.905201
Average	0.924888

Nilai yang dihasilkan untuk menghasilkan solusi *cluster* yang baik harus mendekati dengan nilai 1. Sehingga semakin besar nilai mendekati 1 yang dihasilkan pada suatu metode hirarki dari perhitungan uji performa *cophenetic correlation coefficient*, maka semakin baik pula *cluster* yang dihasilkan. Pada Tabel 4 dapat dilihat pengujian performa dengan menggunakan metode *cophenetic correlation coefficient* menunjukkan nilai uji pada metode average linkage memiliki nilai 0.9248882 yang lebih besar dari metode single linkage senilai 0.905201 dan complete linkage senilai 0.805725. hal ini menunjukkan bahwa metode average linkage baik digunakan dalam membentuk suatu cluster di dalam penelitian ini karena nilai yang dihasilkan mendekati nilai 1. Hal ini dikarenakan metode *average linkage clustering* merupakan satu-satunya metode *clustering*

secara hierarki yang memperhitungkan setiap jarak antar titiknya dalam menentukan urutan membentuk *cluster*. Grafik hasil nilai uji performa menggunakan *cophenetic correlation coefficient* dapat dilihat pada Gambar 4.

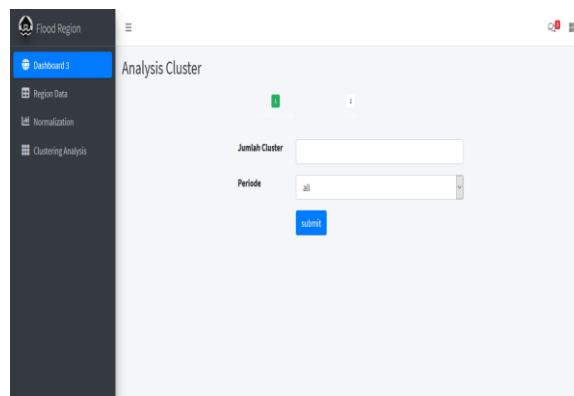


Gambar 4. Cophenetic Correlation Coefficient

Pada Gambar 4 hasil uji metode *cophenetic correlation coefficient* divisualisasikan dalam grafik bar horizontal. dapat dilihat dimana sumbu x adalah nilai uji dengan rentang nilai 0 sampai 1 dimana *cluster* yang terbentuk dari *Hierarchical Clustering* semakin bagus bila nilai *cophenetic* mendekati nilai 1. Sedangkan sumbu y adalah besar nilai pada setiap metode yaitu metode *average linkage*, *single linkage*, *complete linkage*. Apabila dilihat pada grafik bar Gambar 4 terlihat bahwa hasil uji metode *average linkage* memiliki nilai yang paling mendekati nilai 1 yaitu sebesar 0.924888.

### 3.3. Implementasi Aplikasi

Implementasi sistem pada penelitian ini dilakukan dengan menggunakan bahasa pemrograman *python* dengan bantuan *framework flask*. Sistem yang dibangun terdiri dari halaman *clustering analysis* dapat dilihat pada Gambar 5 digunakan untuk melakukan analisis kluster berdasarkan inputan jumlah kluster dan tahun. Halaman untuk menampilkan hasil analisis *cluster* berdasarkan inputan yang dilakukan pada Gambar 5 dapat dilihat pada Gambar 6.



Gambar 5. Analisis Cluster

Pada Gambar 5 merupakan halaman untuk analisis kluster. Pada halaman ini akan ditampilkan *form* untuk melakukan input analisis kluster yang terdiri dari dua inputan yaitu periode tahun yang akan dikluster, inputan

banyak kluster yang akan dibentuk, dan tombol submit untuk proses analisis kluster. Hasil keluaran analisis dapat dilihat pada Gambar 6.

No	tahun	kota	latitude	longitude	riwayat	korban	terendam	cluster
0	2014	Bangkalan	-7.02809	112.747	2.0	125.0	623.0	2
1	2014	Banyuwangi	-8.21029	114.374	1.0	5000.0	0.0	2
2	2014	Bitar	-8.09719	112.342	0.0	0.0	0.0	2
3	2014	Bojonegoro	-7.1538741	111.8818807	8.0	4115.0	986.0	0
4	2014	Bondowoso	-7.88942	113.035	0.0	0.0	0.0	2
5	2014	Gresik	-7.27435	112.31	2.0	0.0	2074.0	2
6	2014	Jember	-8.21893	113.075	6.0	8784.0	3414.0	0
7	2014	Jombang	-7.57409	112.286	1.0	13393.0	0.0	2

Gambar 6. Tampilan hasil cluster

Pada Gambar 6 merupakan halaman menampilkan hasil analisis cluster. Halaman ini berisi berupa tabel yang terdiri dari tahun, kota, latitude, longitude, Riwayat, korban, terendam, cluster. Hasil analisis ini yang akan digunakan sebagai acuan dalam menyajikan informasi dalam sistem informasi geografis.

No	Tahun	Kota	Hujan	Kepadatan	Riwayat Banjir	Mengungsi	Terendam
1	2014	Bangkalan	1805.52	743.0	2.0	125.0	623.0
2	2014	Banyuwangi	99.7	275.0	1.0	5000.0	0.0
3	2014	Bitar	13.98	718.0	0.0	0.0	0.0
4	2014	Bojonegoro	60.79	630.0	8.0	4115.0	986.0
5	2014	Bondowoso	125.29	485.0	0.0	0.0	0.0
6	2014	Gresik	154.5	1108.0	2.0	0.0	2074.0
7	2014	Jember	205.23	725.0	6.0	8784.0	3414.0
8	2014	Jombang	983.0	1065.0	1.0	13393.0	0.0

Gambar 7. Data Wilayah

Pada Gambar 7 merupakan halaman untuk menyimpan data dan menampilkan informasi mengenai data kota, tahun, curah hujan, kepadatan penduduk, Riwayat banjir, korban mengungsi, dan rumah terendam. Pada halaman ini terdiri dari satu inputan upload yang digunakan untuk melakukan import data eksternal, dan sebuah tombol untuk melakukan normalisasi pada data. Sedangkan pada pojok sebelah kanan terdapat satu tombol ekspor yang dapat digunakan untuk melakukan ekspor data pada sistem dalam bentuk file format csv.

#### 4. Kesimpulan

Hasil pemodelan dampak banjir pada wilayah jawa timur menggunakan agglomerative hierarchical clustering dengan metode average linkage menunjukkan hasil akurasi dengan *cophenetic correlation coefficient* sebesar 0.924888 dengan menggunakan 950 data. Aplikasi yang dirancang dengan visualisasi sistem informasi geografis sangat membantu mengidentifikasi wilayah/kota yang memiliki dampak banjir rendah, sedang, tinggi. Namun atribut dalam penelitian ini

tampak masih belum sepenuhnya mendukung untuk kejadian di lapangan sebenarnya sehingga masih perlu untuk mempertimbangkan faktor yang lain terkait bencana banjir agar hasil pemetaan sesuai dengan apa yang terjadi di lapangan.

#### Daftar Pustaka

- [1] C. Hasiholan, R. Primananda, and K. Amron, "Implementasi Konsep Internet Of Things Pada Sistem Monitoring Banjir Menggunakan Protokol MQTT," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 12, pp. 6128–6135, 2018.
- [2] S. Adi, "Characterization of Flash Flood Disaster in Indonesia (Karakterisasi Bencana Banjir Bandang di Indonesia)," *J. Sains dan Teknol. Indones.*, vol. 15, no. 1, pp. 42–51, 2013.
- [3] D. S. Ratnasari and P. Kusumawardani, "Pemetaan Risiko Bencana di Kota Bogor Tahun 2015 ( Bencana Banjir , Tanah Longsor , Angin Puting Beliung , Dan Kebakaran )," *Semin. Nas. Penginderaan Jauh*, vol. 2015, pp. 720–839, 2016.
- [4] B. Liu, *Web Data Mining Exploring Hyperlinks, Contents, and Usage Data*, vol. 11, no. 5. 2015.
- [5] A. Bouguettaya, Q. Yu, X. Liu, X. Zhou, and A. Song, "Efficient Agglomerative Hierarchical Clustering," *Expert Syst. Appl.*, vol. 42, no. 5, pp. 2785–2797, 2015, doi: 10.1016/j.eswa.2014.09.054.
- [6] A. Ramadhan, Mustakim, and R. Handinata, "Implementasi Algoritma Fuzzy C Means Dan Moora Untuk Pengelompokan Dan Penentuan Wilayah Penanggulangan Bencana Banjir," no. November, p. Pekanbaru, 2019.
- [7] K. K. Raihana, S. M. K. Rishad, T. Sadia, S. Ahmed, M. S. Alam, and R. M. Rahman, "Identifying Flood Prone Regions In Bangladesh By Clustering," *Proc. - 17th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2018*, pp. 556–561, 2018, doi: 10.1109/ICIS.2018.8466533.
- [8] L. N. Wang, X. H. Chen, Q. X. Shao, and Y. Li, "Flood Indicators And Their Clustering Features In Wujiang River, South China," *Ecol. Eng.*, vol. 76, pp. 66–74, 2015, doi: 10.1016/j.ecoleng.2014.03.018.
- [9] P. Prihandoko and B. Bertalya, "A Data Analysis Of The Impact Of Natural Disaster Using K-Means Clustering Algorithm," *Kursor*, vol. 8, no. 4, p. 169, 2017, doi: 10.28961/kursor.v8i4.109.
- [10] S. G. K. Patro and K. K. sahu, "Normalization: A Preprocessing Stage," *Iarjset*, pp. 20–22, 2015, doi: 10.17148/iarjset.2015.2305.
- [11] D. Marutho, S. Hendra Handaka, E. Wijaya, and Muljono, "The Determination Of Cluster Number At K-Mean Using Elbow Method And

- Purity Evaluation On Headline News,” *Proc. - 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life, iSemantic 2018*, pp. 533–538, 2018, doi: 10.1109/ISEMANTIC.2018.8549751.
- [12] M. Roux, “A Comparative Study Of Divisive And Agglomerative Hierarchical Clustering Algorithms,” *J. Classif.*, vol. 35, no. 2, pp. 345–366, 2018, doi: 10.1007/s00357-018-9259-9.
- [13] N. Bhargava, A. Kumawat, and R. Bhargava, “Fingerprint Matching of Normalized Image Based On Euclidean Distance,” *Int. J. Comput. Appl.*, vol. 120, no. 24, pp. 20–23, 2015, doi: 10.5120/21409-4428.
- [14] S. Kumar and D. Toshniwal, “Analysis Of Hourly Road Accident Counts Using Hierarchical Clustering And Cophenetic Correlation Coefficient (CPCC),” *J. Big Data*, vol. 3, no. 1, pp. 1–11, 2016, doi: 10.1186/s40537-016-0046-3.
- [15] D. Saptono, T. M. Sampurna, and T. W. R. N, “Implementasi Algoritma Gunning Fog Index Pada Uji Keterbacaan ( Readability Test ) Bahasa Indonesia Menggunakan Bahasa Pemrograman Python,” *Semantik*, vol. 3, no. 1, pp. 72–77, 2013.