

**ZONASI DAERAH TERDAMPAK BENCANA ANGIN PUTING BELIUNG
MENGUNAKAN K-MEANS *CLUSTERING* DENGAN ANALISIS
SILHOUETTE COEFFICIENT, DAVIES BOULDIN INDEX DAN *PURITY***

SKRIPSI



Disusun Oleh

RAVENA LAILATUR ROHMAH

NIM. H02215007

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN AMPEL

SURABAYA

2019

LEMBAR KEASLIAN

Saya yang bertanda tangan dibawah ini,

Nama : Ravena Lailatur Rohmah

NIM : H02215007

Program Studi : Matematika

Angkatan : 2015

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul: **“ZONASI DAERAH TERDAMPAK BENCANA ANGIN PUTING BELIUNG MENGGUNAKAN K-MEANS CLUSTERING DENGAN ANALISIS SILHOUETTE COEFFICIENT, DAVIES BOULDIN INDEX DAN PURITY”**. Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 31 Juli 2019

Yang Menyatakan



Ravena Lailatur Rohmah

NIM. H02215007

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

NAMA : Ravena Lailatur Rohmah

NIM : H02215007

JUDUL : Zonasi Daerah Terdampak Bencana Angin Puting Beliung
Menggunakan K-Means *Clustering* dengan Analisis *Silhouette Coefficient, Davies Bouldin Index* dan *Purity*

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 31 Juli 2019

Dosen Pembimbing I



(Dian C. Rini Novitasari, M. Kom)
NIP. 198511242014032001

Dosen Pembimbing II



(Wika Dianita Utami, M.Sc)
NIP. 199206102018012003

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

NAMA : Ravena Lailatur Rohmah

NIM : H02215007

JUDUL : Zonasi Daerah Terdampak Bencana Angin Puting Beliung Menggunakan K-Means *Clustering* dengan Analisis *Silhouette Coefficient*, *Davies Bouldin Index* dan *Purity*

Telah dipertahankan di depan tim penguji skripsi

Pada hari Rabu, 31 Juli 2019

Mengesahkan,

Tim Penguji

Penguji I



(Dian C. Rini Novitasari, M. Kom)
NIP. 198511242014032001

Penguji II



(Wika Dianita Utami, M.Sc)
NIP. 199206102018012003

Penguji III



(Nurissaidah Ulinuha, M.Kom)
NIP. 1999011022014032004

Penguji IV



(Putrode Keumala Intan, M.Si)
NIP. 198805282018012001

Mengetahui,

Dekan Fakultas Sains dan Teknologi

Sunan Ampel Surabaya




(Eni Purwanti, M.Ag)
NIP. 196512211990022001



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : Ravena Lailatur Rohmah
NIM : H02216007
Fakultas/Jurusan : Sains dan Teknologi / Matematika
E-mail address : lailatur.ravena@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Skripsi Tesis Desertasi Lain-lain (.....)

yang berjudul :

Zonasi Daerah Terdampak Bencana Angin Puting
Beliung Menggunakan k-Means Clustering dengan
Analisis Silhouette Coefficient, Davies-Bouldin Index dan Purity

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

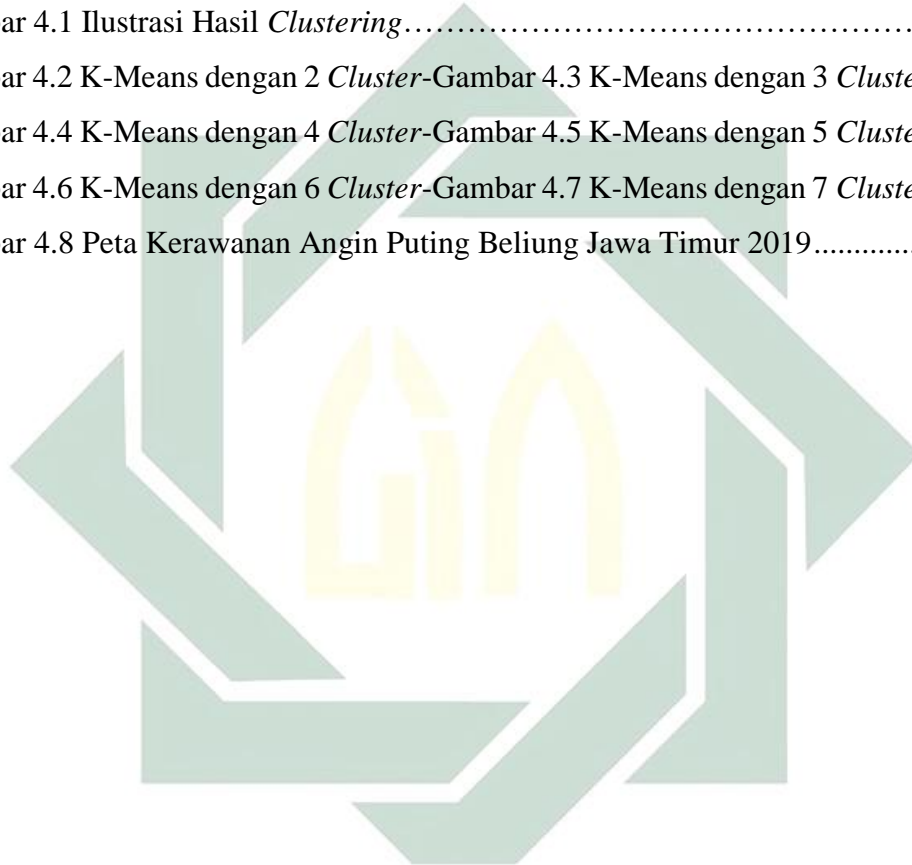
Surabaya, 31 Juli 2019

Penulis

(Ravena Lailatur R.)
nama terang dan tanda tangan

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Penentuan Keanggotaan Kelompok Berdasarkan Jarak.....	14
Gambar 2. 2 Ilustrasi Grafik <i>Silhouette Coefficient</i>	18
Gambar 4.1 Ilustrasi Hasil <i>Clustering</i>	44
Gambar 4.2 K-Means dengan 2 <i>Cluster</i> -Gambar 4.3 K-Means dengan 3 <i>Cluster</i>	49
Gambar 4.4 K-Means dengan 4 <i>Cluster</i> -Gambar 4.5 K-Means dengan 5 <i>Cluster</i>	50
Gambar 4.6 K-Means dengan 6 <i>Cluster</i> -Gambar 4.7 K-Means dengan 7 <i>Cluster</i>	50
Gambar 4.8 Peta Kerawanan Angin Puting Beliung Jawa Timur 2019.....	57



DAFTAR TABEL

Tabel 2.1 Tabel Nilai <i>Silhouette</i> Kaufman dan Rousseeuw	17
Tabel 3.1 Definisi Operasional Variabel.....	31
Tabel 4. 1 Data Kabupaten dan Kota Terdampak Angin Puting Beliung di Jawa Timur 2019.....	36
Tabel 4.2 Centroid iterasi Ke-1	35
Tabel 4.3 Hasil Perhitungan Jarak Euclidean Iterasi 1	38
Tabel 4.4 <i>Centroid Cluster</i> 1.....	38
Tabel 4. 5 <i>Centroid Cluster</i> 2.....	39
Tabel 4.6 <i>Centroid</i> untuk Iterasi ke-2	39
Tabel 4. 7 Iterasi ke-2.....	39
Tabel 4.8 <i>Centroid</i> Terbaru.....	40
Tabel 4.9 Iterasi ke-3.....	40
Tabel 4.10 Hasil Cluster 1.....	41
Tabel 4.11 Hasil Cluster 2.....	41
Tabel 4.12 Hasil <i>Cluster</i> 3	41
Tabel 4.13 Hasil Cluster 4.....	42
Tabel 4.14 Hasil Contoh Perhitungan Nilai <i>Silhouette Coefficient</i>	44
Tabel 4. 15 Titik pusat <i>cluster</i> akhir dari proses <i>clustering</i>	46
Tabel 4.16 Nilai <i>Sum of Square Between Cluster</i>	46
Tabel 4.17 Nilai Rasio Maksimum	47
Tabel 4. 18 Hasil <i>Cluster</i> Kota/Kab. di Jawa Timur Tahun 2019.....	48
Tabel 4.19 Tabel Uji Validasi <i>cluster</i>	49
Tabel 4. 20 Centroid Akhir Pembentukan <i>Cluster</i> Menggunakan K-Means	52
Tabel 4. 21 Pola Pada <i>Centroid</i> Akhir	53

puting beliung merupakan bencana paling sering terjadi di Indonesia yaitu sebanyak 698 kali per Juni 2019 (diambil dari website <http://dibi.bnpb.go.id/>). Angin puting beliung yaitu angin yang berputar sangat cepat yang membenteng seperti badai dan memiliki titik pusat. Faktor-faktor penyebab terjadinya angin puting beliung tidak tetap seperti bencana yang lain, hal itu menyebabkan sulitnya memprediksi bencana tersebut (Triana, Hadi, and Husain 2017). Penduduk yang tinggal di wilayah yang rawan bencana angin puting beliung harus selalu siaga mengingat bencana tersebut dapat terjadi kapan saja.

Buku Rencana Aksi Nasional Pengurangan Risiko Bencana (RAN-PRB) 2006-2009 menyebutkan bahwa kejadian bencana di Indonesia terus meningkat dari tahun ke tahun. Bencana-bencana tersebut menyebabkan kerusakan yang besar dalam kehidupan manusia dan sekitarnya, diantaranya adalah infrastruktur ekonomi dan sosial, serta kerusakan lingkungan yang lain (Yana et al. 2018). Diantaranya korban meninggal, kerusakan rumah, kerusakan fasilitas umum, dan lain sebagainya.

Badan Nasional Penanggulangan Bencana (BNPB) merupakan salah satu lembaga pemerintah yang menjadi pelaksana penanggulangan bencana di Indonesia. BNPB memerlukan data yang tepat dan akurat secara cepat ketika terjadi bencana. Hal ini diperlukan agar dapat terbentuk penanggulangan yang tepat dan koordinasi yang baik antar dinas terkait (Yana et al. 2018). Untuk merealisasikan hal tersebut dibutuhkan suatu alat atau analisis data yang dapat membantu proses pengumpulan informasi. Salah satunya yaitu merealisasikan penelitian yang akan dilakukan oleh penulis.

Dalam Al-Quran Surat Al Ahqof ayat 24-25 yang artinya, “Maka tatkala mereka melihat adzab itu berupa awan yang menuju ke lembah-lembah mereka, berkatalah mereka, “Inilah awan yang akan menurunkan hujan kepada kami”.(Bukan!) bahkan Itulah azab yang kamu minta supaya datang dengan segera (yaitu) angin yang mengandung azab yang pedih, yang menghancurkan segala sesuatu dengan perintah Tuhannya. Maka jadilah mereka tidak ada yang kelihatan lagi kecuali (bekas-bekas) tempat tinggal mereka. Demikianlah Kami memberi balasan kepada kaum yang berdosa”. Firman Allah Q.S Al-Ahqof 24-25 dapat dimaknai bahwa tidak ada yang kelihatan lagi kecuali (bekas-bekas) tempat tinggal mereka adalah tidak ada lagi satu pun orang yang hidup di dalam rumah-rumah mereka. Artinya seluruh mereka hancur dan mati disebabkan angin tersebut cukup dalam sekejap mata saja dengan sekali hembusan. Sungguh ini adalah tanda dan bukti kekuasaan Allah yang sangat besar. Seyogyanya orang-orang yang beriman mengambil pelajaran dariNya.

Mitigasi bencana merupakan upaya yang diperlukan guna mengurangi dampak yang terjadi akibat suatu bencana termasuk kesiapan dalam menghadapi resiko jangka panjang. Penelitian kebencanaan juga telah dilakukan (Sari 2018) dengan menyatakan bahwa pengelompokan terhadap wilayah-wilayah di Indonesia yang mengalami gempa bumi. Hal itu dilakukan agar dapat membantu pemerintah dalam melakukan upaya mitigasi terutama terhadap wilayah yang rawan gempa bumi. Maka dari itu perlu dilakukan pula pengelompokan daerah rawan bencana alam lainnya.

Menurut penelitian yang telah dilakukan Vincent, dkk. mencari dan mengabadikan moment bencana angin puting beliung tidak mudah karena kejadian tersebut berangsur sangat cepat. Dalam penelitian lain, telah diprediksi terjadinya angin puting beliung di Kanada menggunakan pendekatan pemodelan Bayesian (Cheng et al. 2013). Florent Renard juga melakukan penelitian dengan tujuan analisis kerentanan fisik teritorial dalam menghadapi banjir menggunakan pengetahuan spasial yang tepat. Hasil dari penelitian ini ditranskripsikan dalam bentuk kartografi oleh GIS menggunakan perawatan geomatika dan grid (Renard 2017).

Penelitian dengan permasalahan bencana angin puting beliung diantaranya yaitu penelitian yang dilakukan oleh Marselina pada tahun 2015 yang berjudul “Analisis Statistika terhadap Penyebab Puting Beliung di Daerah Istimewa Yogyakarta 2011-2014”. Penelitian tersebut menggunakan variabel prediktor curah hujan, suhu udara, tekanan udara, kelembaban udara, arah angin, kecepatan angin, dan ketinggian daerah. Penelitian tersebut menggunakan metode regresi logistik. Kesimpulan dari penelitiannya yaitu kejadian angin puting beliung memiliki variabel prediktor signifikan yang berbeda-beda bergantung pada posisi masing-masing wilayah terhadap ketinggiannya dari permukaan laut. Penelitian pemetaan daerah rawan longsor yang telah dilakukan oleh M. Leng dkk, disebutkan bahwa tingkat kerawanan longsor dibagi kedalam tiga *cluster* Kabupaten Timor Tengah Utar. Penelitian ini menggunakan GIS untuk mengetahui daerah rawan longsor tersebut dimana menunjukkan bahwa GIS berperan sangat penting dalam upaya

meningkatkan mitigasi bencana daerah. Pada penelitian ini GIS berperan penting dalam pewarnaan zonasi dengan karakter daerah rawan, sangat rawan, dan aman.

Deng, dkk (2016) juga telah melakukan penelitian dengan judul “*A Few Clarification on Tornado Density Mapping*” yang menjelaskan tentang seberapa besar faktor GIS berpengaruh terhadap proses pemetaan kepadatan angin puting beliung. Tanesib, dkk (2018) telah melakukan penelitian kebencanaan menggunakan GIS. Pada penelitian tersebut terbukti mampu menyediakan sistem analisa keruangan yang akurat serta menyediakan informasi data geospasial setiap objek di permukaan bumi secara cepat sehingga dapat dilakukan upaya mitigasi yang bertujuan untuk mencegah bahaya (resiko) yang berpotensi menjadi bencana atau mengurangi efek dari bencana ketika bencana tersebut sudah terjadi.

Penerapan ilmu data mining dalam kehidupan sehari-hari sangatlah luas. Apalagi dengan adanya revolusi industri 4.0 yang selalu melibatkan teknologi dalam berbagai bidang, hal ini mendukung pemanfaatan data mining. Seperti dalam bidang pertanian yang telah dilakukan oleh (Afifah, Rini, and Lubab 2016) dengan judul “Pengklateran Lahan Sawah di Indonesia Sebagai Evaluasi Ketersediaan Produksi Pangan Menggunakan *Fuzzy C-Means*” dengan tujuan mengetahui kelompok luas lahan sawah di Indonesia dengan tiga *cluster* yaitu luas, sedang, sempit. Dalam bidang bisnis juga telah dilakukan penerapan metode K-Means oleh (Murti 2017) dengan judul “Penerapan Metode K-Means Untuk Mengelompokkan Potensi Produksi Buah-Buahan di Provinsi Daerah Istimewa Yogyakarta” dengan tujuan untuk mengetahui manakah daerah yang menghasilkan buah sedikit, sedang, dan terbanyak. Pada bidang kesehatan, data mining bermanfaat untuk mengetahui

tingkat kebutuhan transfusi darah dari penderita Thalassaemia, hal itu dilakukan untuk mendapatkan informasi seperti penentuan kriteria suatu penyakit (Sulastri and Gufroni 2017). Salah satu pemanfaatan data mining dalam bidang pendidikan adalah menggunakan algoritma K-Means dalam melakukan pencarian calon mahasiswa baru dari berbagai kota di Indonesia, hal ini dilakukan oleh pihak marketing President University (Ong 2013).

Yana et al. (2018) telah melakukan bahwa K-Means merupakan salah satu metode *clustering* yang sederhana namun memiliki keakuratan yang tinggi. Metode tersebut bertujuan untuk mempartisi data ke dalam dua atau lebih kelompok. Penelitian tersebut bertujuan untuk mengelompokkan wilayah dengan karakteristik jumlah kejadian bencana *relative* sama sehingga dapat diketahui wilayah mana saja yang rawan terjadi bencana alam.

Metode K-Means juga telah digunakan dalam penelitian (Sirat, Setiawan, and Ramdani 2018) dengan judul “Analisis Perbandingan Algoritma K-Means dan Isodata untuk klasterisasi Data Kejadian Titik Api di Wilayah Sumatera Pada Tahun 2001 hingga 2014”. Penelitian tersebut, disebutkan bahwa metode K-Means memiliki keunggulan dibandingkan algoritma isodata. Selain itu, penelitian tersebut juga menyebutkan bahwa nilai *silhouette coefficient* (SC) pada algoritma K-Means memiliki nilai yang lebih tinggi dibandingkan algoritma isodata. Penelitian lain menyebutkan bahwa hasil *clustering* menggunakan uji validasi *Silhouette*, *Davies Bouldin* dan *Calinski Harabasz* memiliki hasil yang lebih baik (Khairati et al. 2019). Penelitian lain juga telah dilakukan oleh (Dehotman 2018) dengan judul “Peningkatan Hasil Evaluasi Clustering Davies Bouldin Index dengan Penentuan

Titik Pusat *Cluster* Awal Algoritma K-Means”. Maka dari itu, pada penelitian ini akan dilakukan uji validasi *cluster* sebanyak tiga metode untuk mengetahui setiap pola metode uji validasi *cluster* tersebut. Darman (2019) telah melakukan penelitian dengan judul “Analisis Data Kejadian Bencana Angin Puting Beliung dengan Metode Online Analytical Processing (OLAP)” dimana dalam penelitian tersebut menghasilkan bahwa Pulau Jawa merupakan daerah yang memiliki kerawanan angin puting beliung yang cukup tinggi.

Dari uraian di atas, peneliti ingin melakukan penelitian yang bertujuan untuk mengelompokkan wilayah atau daerah yang rawan bencana angin puting beliung. Metode yang akan digunakan dalam pengelompokan tersebut adalah K-Means yang dianalisis dengan *Silhouette Coefficient*, *Davies Bouldin Index* dan *Purity*. Pemilihan metode ini didasarkan oleh penelitian yang telah dilakukan (Yana, dkk, 2018) yang menyebutkan bahwa salah satu metode *clustering* yang simpel, sederhana dan memiliki nilai optimal *cluster* yang tinggi adalah metode K-Means. Analisis *cluster* menggunakan *silhouette coefficient*, *davies bouldin index* dan *purity* diterapkan karena peneliti ingin mengetahui apakah *clustering* dengan pemilihan k yang tepat dan optimal serta struktur data yang kuat. Hal ini juga dilakukan dalam penelitian (Swindiarito, 2018) bahwa analisis *silhouette* dilakukan dalam *clustering* untuk melihat kekuatan objek dari data yang terbentuk. Dimana, analisis kekuatan struktur data pada percobaan sejumlah k *cluster* sangat diperlukan untuk mengetahui dan menganalisis ada atau tidaknya perpindahan data dari *cluster* satu ke *cluster* yang lainnya.

Pemilihan analisis *cluster* juga didasarkan oleh penelitian yang telah dilakukan oleh (Sari 2018) bahwa dari beberapa analisis *cluster* dapat mengetahui berapa jumlah *cluster* optimal. Dalam penelitiannya ia menggunakan, *C-Index*, *Davies Bouldin Index*, *McClain Rao Index*, *Silhouette Index*, dan *Connectivity Index* didapatkan jumlah *cluster* yang terbentuk yaitu sebanyak 2 *cluster* dan 4 *cluster*. Dengan menggunakan *cluster variance* didapatkan bahwa *cluster* paling baik jika menggunakan 4 *cluster* karena memiliki nilai rasio S_w dan S_b yang terkecil. Pada penelitian ini juga dilakukan representasi dari hasil *cluster* menggunakan pewarnaan GIS, seperti yang telah diuraikan oleh (Hardaningrum, Sulaeman, & Supriyana, 2016) dalam penelitiannya tentang zonasi rawan bencana gempa bumi kota Malang, bahwa zonasi rawan bencana alam sangatlah penting karena hal itu merupakan salah satu usaha meminimalkan kerugian harta benda maupun jiwa yang disebabkan oleh bencana alam. Untuk mempermudah pemerintah mengetahui daerah-daerah rawan angin puting beliung maka setelah terbentuk *cluster* yang tepat maka dilakukan pewarnaan wilayah ke dalam peta daerah rawan bencana angin puting beliung menggunakan *Geographic Information System* (GIS).

Berdasarkan pewarnaan GIS tersebut maka dapat teridentifikasi wilayah (kota atau kabupaten) mana saja yang memiliki intensitas tinggi terkena bencana angin puting beliung. Hal ini diupayakan karena merupakan salah satu penerapan mitigasi bencana, artinya penelitian dilakukan untuk menjadi salah satu acuan pemerintah dalam mempersiapkan kewaspadaan terhadap bencana angin puting beliung. Penelitian ini mengacu pada data BNPB berdasarkan wilayah sekitar yang terdampak akibat bencana angin puting beliung di Jawa Timur, maka peneliti akan

Kemudian dihitung jarak antara setiap pusat *cluster* dengan setiap data. Untuk mengetahui jarak tersebut digunakan rumus Euclidean, seperti pada Persamaan (2.1) (Murti 2017).

Berikut tahapan-tahapan yang dilakukan dalam mengelompokkan data dengan menggunakan metode K-Means (Everitt et al. 2011).

1. Menentukan banyaknya k , dimana k merupakan jumlah *cluster* yang akan dibentuk.
2. Menentukan titik pusat atau biasa disebut *centroid* dari setiap *cluster*. Pengambilan *centroid* ini secara acak.
3. Alokasikan semua data ke centroid terdekat, Untuk mengalokasikan tersebut yaitu dengan menghitung jarak setiap objek atau data ke titik pusat (*centroid*). Jarak *centroid* yang digunakan yaitu *euclidean* dengan Persamaan 2.1 berikut.

$$d(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (2.1)$$

Dimana :

$d(x, y)$ = jarak data ke x ke pusat klaster y

x_i = data x pada observasi ke- i

y_i = titik pusat ke y observasi ke- i

n = banyaknya observasi

4. Mengelompokkan setiap data berdasarkan kedekatannya (jarak terkecil) dengan *centroid* menggunakan Persamaan 2.2 berikut.

$$\text{Min } \sum_{k=1}^k d(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} \quad (2.2)$$

5. Mencari nilai titik pusat atau *centroid* terbaru. Nilai tersebut diperoleh dari rata-rata *cluster* setiap *cluster* dengan menggunakan Persamaan 2.3 berikut.

$$C_{ij} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (2.3)$$

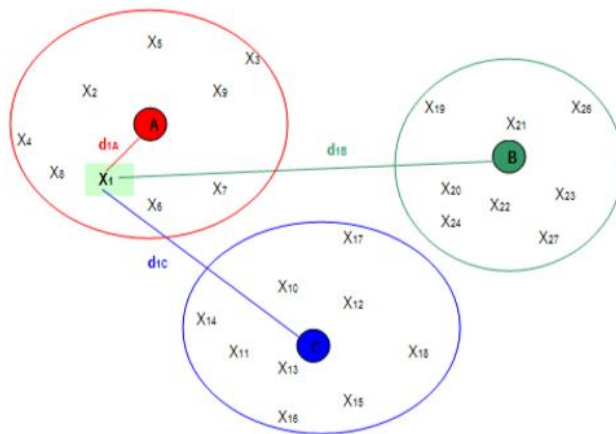
Dimana :

C_{ij} = *Centroid* terbaru pada iterasi k

x_{ij} = anggota *cluster* ke k

p = Banyaknya anggota *cluster* ke k

6. Melakukan perulangan langkah 2-5 hingga kondisi konvergen tercapai, yaitu anggota tiap *cluster* tidak mengalami perubahan letak *cluster*.



Gambar 2.1 Ilustrasi Penentuan Keanggotaan Kelompok Berdasarkan Jarak

Sumber Gambar: (Prilianti and Kunci 2014)

Gambar 2.1 merupakan ilustrasi penentuan keanggotaan kelompok berdasarkan jarak *euclidean*. Huruf A, B, C yang berada di tengah lingkaran merupakan titik pusat

atau *centroid* dimana d_{1A} merupakan jarak data x_1 terhadap titik pusat *cluster* a. Pada Gambar 2.1 dilakukan *cluster* dengan jumlah *cluster* sebanyak 3.

2.2 Uji Validasi Cluster

Setelah dilakukan *clustering*, maka perlu dilakukan analisis *cluster*. Analisis *cluster* merupakan salah satu langkah dari uji validasi atau evaluasi *cluster*. Validasi *cluster* bertujuan untuk mengambil keputusan dari hasil *cluster* secara objektif dan kuantitatif. Menurut (Gardon 1999) terdapat beberapa pendekatan utama untuk validasi *cluster* yaitu:

2.2.1 External test,

Membagi data menjadi dua bagian dimana data hasil klasifikasi dimasukkan dan kemudian dibandingkan dengan data hasil klasifikasi yang tidak termasuk dalam klasifikasi dasar.

2.2.2 Internal test,

Proses yang menghitung hubungan antara struktur *cluster* yang diperoleh dengan data itu sendiri. Data yang digunakan untuk proses *cluster* dan validasi sama.

2.2.3 Relative test,

Membandingkan hasil klasifikasi yang berbeda dari satu kumpulan data dengan menggunakan klasifikasi yang sama dengan parameter yang berbeda.

2.3 Silhouette Coefficient

Silhouette coefficient merupakan metode evaluasi untuk menguji optimal atau ketepatan sebuah *cluster* yang telah terbentuk dari proses *clustering* (Tanzil Furqon and Muflikhah 2016). *Silhouette coefficient* memberikan hasil kualitas visual objek dalam tiap *cluster*, memberikan informasi sesuai dengan jumlah *cluster* pada data set. Untuk setiap objek dinotasikan oleh *cluster* dimana dia berasal (Swindiarto 2018). Metode ini merupakan gabungan dari metode *separation* dan *cohesion*. Tahapan perhitungan *Silhouette coefficient* adalah sebagai berikut (Handoyo, Rumani, and Nasution, 2014):

1. Hitung rata-rata jarak dari suatu data, menggunakan Persamaan 2.4 maka didapatkan rata-rata dengan cara memisalkan i terhadap semua data lain yang berada dalam satu *cluster* sebagai berikut.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (2.4)$$

dimana :

$a(i)$ = Perbedaan rata-rata objek (i) ke semua objek lain pada A

$d(i, j)$ = jarak antara data i dengan j

A = Cluster

2. Hitung rata-rata jarak data i tersebut dengan semua data di *cluster* lain, dan diambil nilai terkecilnya menggunakan Persamaan 2.5.

$$d(i,C) = \frac{1}{|C|} \sum_{j \in C} d(i,j) \quad (2.5)$$

dimana :

$d(i,C)$ = Perbedaan rata-rata objek (i) ke semua objek lain pada C

C = *cluster* lain selain *cluster* A atau *cluster* C tidak sama dengan *cluster* A.

3. Setelah menghitung $d(i,C)$ untuk semua C, maka diambil nilai terkecil dengan menggunakan Persamaan 2.6.

$$b(i) = \min_{C \neq A} d(i,C) \quad (2.6)$$

Cluster B yang mencapai minimum (yaitu, $d(i,B) = b(i)$) disebut tetangga dari objek (i). Ini adalah *cluster* terbaik kedua untuk objek (i).

4. Nilai *Silhouette Coefficient* didefinisikan seperti pada Persamaan 2.7.

$$s(i) = \frac{b(i) - a(i)}{\max a(i), b(i)} \quad (2.7)$$

Tabel 2.1 Tabel Nilai *Silhouette* Kaufman dan Rousseeuw

Nilai <i>Silhouette Coefficient</i>	Struktur
$0.7 < SC \leq 1$	Struktur Kuat
$0.5 < SC \leq 0.7$	Struktur Sedang
$0.25 < SC \leq 0.5$	Struktur Lemah
$SC \leq 0.25$	Tidak terstruktur

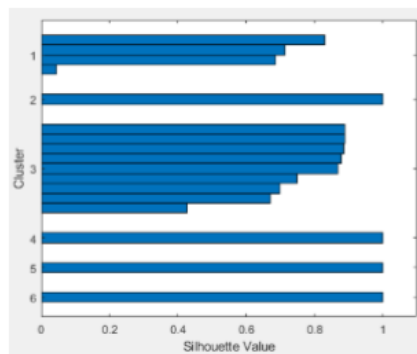


Figure 3. FCM with 6 Clusters (C = 6)

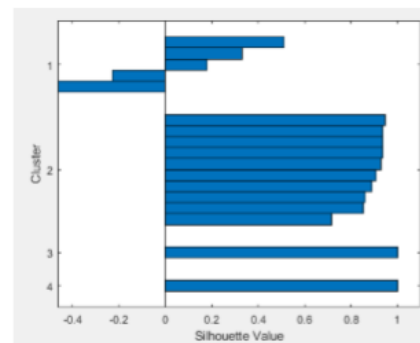


Figure 5. FCM with 4 Clusters (C = 4)

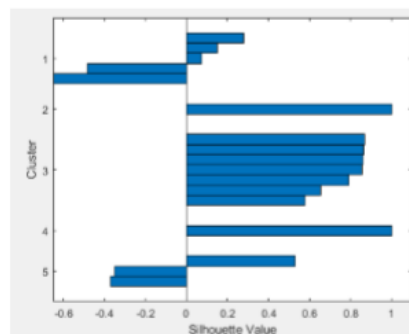


Figure 4. FCM with 5 Clusters (C = 5)

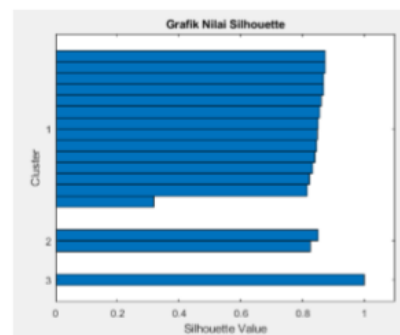


Figure 6. FCM with 3 Clusters (C = 3)

Gambar 2. 2 Ilustrasi Grafik *Silhouette Coefficient*

Sumber : (Swindiarto 2018)

Gambar 2.2 merupakan ilustrasi hasil *silhouette coefficient* dimana dari grafik dalam gambar tersebut merepresentasikan data yang terbentuk dalam *cluster*. Jika data tersebut terletak di titik lebih dari sama dengan nol artinya data tersebut tepat pada golongan *cluster* tersebut. Namun, jika data terletak di titik negatif atau kurang dari nol maka data tersebut tidak memiliki struktur yang tepat atau tidak masuk ke dalam *cluster* manapun.

mengukur perbedaan data-data yang ada pada *cluster* yang berbeda. Pengukuran menggunakan DBI memaksimalkan jarak *inter-cluster* antara *cluster* c_j dan c_i dan pada waktu yang sama mencoba untuk meminimalkan jarak antar titik dalam sebuah *cluster*. Jika jarak *intra-cluster* minimal berarti masing-masing objek dalam *cluster* akan terlihat lebih jelas (Meylimdra Arini Permatadevi, Rully A. Hendrawan, S.Kom 2013). Langkah-langkah perhitungan *Davies-Bouldin Index* adalah sebagai berikut:

1. *Sum of Square Within-Cluster (SSW)*

Jika ingin mengetahui kohesi dalam sebuah *cluster* ke- i salah satu caranya yaitu dengan menghitung nilai dari *Sum of Square Within-Cluster (SSW)*. Kohesi diartikan sebagai jumlah dari kedekatan atau kemiripan data terhadap titik pusat *cluster* dari sebuah *cluster* yang diikuti. Persamaan yang digunakan untuk memperoleh nilai *Sum of Square Within-Cluster (SSW)* menggunakan Persamaan 2.8 berikut.

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (2.8)$$

Dimana,

m_i = jumlah data dalam *cluster* ke- i

c_i = *centroid cluster* ke- i

$d(x_j, c_i)$ = jarak setiap data ke *centroid* i yang dihitung menggunakan jarak *euclidean*.

2. *Sum of Square Between-Cluster (SSB)*

Perhitungan *Sum of Square Between-Cluster* (SSB) bertujuan untuk mengetahui separasi atau jarak antar *cluster*. Berikut Persamaan yang digunakan untuk menghitung nilai *Sum of Square Between-Cluster* (SSB):

$$SSB_{i,j} = d(x_i, x_j) \quad (2.9)$$

Dimana, $d(x_i, x_j)$ merupakan jarak antara data ke i dengan data ke j di *cluster* lain.

3. *Ratio* (Rasio)

Cluster yang baik adalah *cluster* yang memiliki nilai kohesi sekecil mungkin dan separasi yang sebesar mungkin. Perhitungan rasio ($R_{i,j}$) ini bertujuan untuk mengetahui nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j untuk menghitung nilai rasio yang dimiliki oleh masing-masing *cluster*. Indeks i dan j merupakan merepresentasikan jumlah *cluster*, dimana jika terdapat 4 *cluster* maka terdapat indeks sebanyak 4 yaitu $i, j, k,$ dan l . Untuk menentukan rasio tersebut digunakan Persamaan 2.10.

$$R_{i,j,\dots,n} = \frac{SSW_i + SSW_j + \dots + SSW_n}{SSB_{i,j} + \dots + SSB_{ni,nj}} \quad (2.10)$$

dimana,

SSW_i = *Sum of Square Within Cluster* pada *centroid* i

$SSB_{i,j}$ = *Sum of Square Between Cluster* data ke i dengan j
pada *cluster* yang berbeda

Pada Persamaan 2.10 n akan berlanjut sejumlah *cluster* yang dipilih dengan syarat ni tidak sama dengan nj .

Bencana merupakan peristiwa dahsyat yang dapat merugikan masyarakat, infrastruktur, dan lingkungan sekitarnya yang melebihi kemampuan masyarakat untuk mengatasi sendiri. Bencana alam sangat berdampak negatif terhadap lingkungan. Oleh karena itu, mengingat tiga komponen utama pembangunan berkelanjutan yaitu lingkungan, masyarakat, dan ekonomi yang terganggu karena bencana maka dibutuhkan sebuah persiapan dan pemetaan daerah rawan bencana (C, Dr, & Dr, 2012).

Menurut BNPB, angin puting beliung adalah angin kencang yang datang secara tiba-tiba, memiliki pusat, bergerak melingkar menyerupai spiral dengan kecepatan 40-50 km/jam hingga menyentuh permukaan bumi serta akan hilang dalam waktu singkat yaitu 3-5 menit (Badan Nasional Penanggulangan Bencana, 2019).

Angin Puting Beliung merupakan fenomena cuaca berasal dari salah satu sumber, yaitu Cumulonimbus (Cb) yang sangat kuat. Tidak semua fenomena yang berasal dari awan Cb ini dapat menjadi angin puting beliung, terkadang hanya berupa hujan lebat yang disertai petir atau hujan es. Masyarakat menyebut bahwa angin puting beliung adalah angin yang kencang yang berputar (*vortex*), dan umumnya terjadi bersamaan dengan curah hujan yang intensitas tinggi. Fenomena ini bersifat lokal, mencakup area antara 5-10 kilometer (Yuniarti, Mazna, 2018).

Puting beliung juga diartikan sebagai angin kencang yang muncul secara tiba-tiba disertai dengan hujan yang memiliki pusat, berintensitas tinggi, bergerak melingkar seperti spiral hingga menyentuh permukaan bumi. Periodenya cukup singkat, berkisar 3-5 menit, mulai dari tumbuh hingga punahnya. Beberapa daerah

menyebut angin ini dengan istilah lesus (jawa), angin puyuh, sirit batara (sunda) (Saputra & Fahrizal, 2017).

BNPB mencatat dan mengumpulkan data wilayah terdampak dari bencana alam yang telah terjadi, data tersebut di publikasikan melalui website resmi milik BNPB yaitu Data Informasi Bencana Indonesia (DIBI). Berikut merupakan dampak yang diakibatkan oleh bencana alam yang telah tercatat oleh BNPB :

2.6.1 Korban Jiwa

Orang/sekelompok orang yang mengalami dampak buruk akibat bencana, seperti kerusakan dana atau kerugian harta benda, penderitaan dan atau kehilangan jiwa. Korban dapat dipilah berdasarkan klasifikasi korban meninggal, hilang, luka atau sakit, menderita dan mengungsi.

- **Meninggal**

Orang yang dilaporkan tewas atau meninggal dunia akibat bencana.

- **Luka-luka**

Orang yang mengalami luka-luka atau sakit, dalam keadaan luka ringan, maupun luka parah/berat, baik yang berobat jalan maupun rawat inap.

- **Hilang**

Orang yang dilaporkan hilang atau tidak ditemukan dan tidak diketahui keberadaannya setelah terjadi bencana.

- **Terdampak**

Orang atau sekelompok orang yang menderita akibat dampak buruk bencana, seperti kerusakan atau kerugian harta benda, namun masih dapat menempati tempat tinggalnya.

- **Mengungsi**

Orang yang terpaksa atau dipaksa keluar dari tempat tinggalnya ke tempat yang lebih aman dalam upaya menyelamatkan diri atau jiwa untuk jangka waktu yang belum pasti sebagai akibat dampak buruk bencana.

2.6.2 Rumah Rusak (Unit)

- **Rusak Berat**

Kriteria kerusakan yang mengakibatkan bangunan roboh atau sebagian besar komponen struktur rusak, sebagai contoh :

- a) Bangunan roboh total/sebagian besar struktur utama bangunan rusak.
- b) Sebagian besar dinding dan lantai bangunan bendung atau dam patah.
- c) Sebagian besar tanggul jebol atau putus.
- d) Saluran pengairan tidak dapat berfungsi.

- **Rusak Sedang**

Kriteria kerusakan yang mengakibatkan sebagian kecil komponen struktur rusak dan komponen penunjang rusak namun bangunan masih tetap berdiri, sebagai contoh :

- a) sebagian kecil struktur utama bangunan rusak.

b) Sebagian besar pintu-pintu air dan komponen penunjang lainnya rusak.

c) Saluran pengairan terputus.

- Rusak Ringan

Kriteria kerusakan yang mengakibatkan sebagian komponen struktur retak (struktur masih bisa digunakan) dan bangunan masih tetap berdiri, sebagian contoh:

a) Sebagian kecil struktur bangunan rusak ringan.

b) Retak-retak pada dinding plesteran.

c) Sebagian kecil pintu-pintu air dan komponen penunjang lainnya rusak.

d) Saluran pengairan masih bisa digunakan.

2.6.3 Fasilitas (Unit)

Kerusakan harta benda meliputi pendidikan (sekolah, madrasah atau pesantren), fasilitas kesehatan (rumah sakit, puskesmas, puskesmas pembantu/pustu), fasilitas peribadatan (masjid, gereja, vihara dan pura), bangunan lain (kantor, pasar, kios) dan jalan yang mengalami kerusakan (rusak ringan, sedang dan berat atau hancur bahkan roboh) serta sawah yang terkena bencana dan puso (gagal panen).

- Fasilitas Kesehatan
- Fasilitas Peribadatan
- Fasilitas Pendidikan

BAB III

METODE PENELITIAN

3.1 Sumber Data

Hasil pencatatan BNPB Indonesia tahun 2018 membuktikan bahwa sejak 2016 sampai dengan awal tahun 2018, sudah lebih dari 2700 bencana alam terjadi di seluruh wilayah yang ada di Indonesia. Berdasarkan data tersebut, pemerintah harus lebih giat dalam mensosialisasikan upaya pengurangan risiko bencana kepada masyarakat. Selain itu, pemerintah juga perlu mengetahui zona-zona atau daerah yang termasuk rawan bencana alam. Data pada ilustrasi Gambar 3.1 menunjukkan angin puting beliung merupakan bencana alam yang paling sering terjadi hingga bulan Juni 2019.

	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Kode_Kab	Nama_Kab/Kot	Kejadian	Nama_Kej	Meninggal	Terluka	Hilang	Menderita	Mengungsi	Rumah_Ri	Rumah_Ri	Rumah_Ri	Kesehatan	Penbadat	Pendidik	usak
2	3579	BATU	PUTING BEL	Puting Beliung	0	0	0	0	0	0	0	4	0	0	0	0
3	3509	JEMBER	PUTING BEL	Puting Beliung	0	0	0	0	5	1	0	0	0	0	0	0
4	3512	SITUBONDO	PUTING BEL	puting beliung	0	0	0	1	0	1	0	0	0	0	0	0
5	3573	MALANG	PUTING BEL	puting beliung	0	0	0	0	0	0	0	1	0	0	0	0
6	3507	MALANG	PUTING BEL	Puting Beliung	0	0	0	0	0	0	0	8	0	0	0	0
7	3509	JEMBER	PUTING BEL	Puting Beliung	0	0	0	0	0	0	0	2	0	0	0	0
8	3504	TULUNGAGUNG	PUTING BEL	Puting Beliung	0	0	0	0	0	0	0	31	0	0	0	0
9	3518	NGANJUK	PUTING BEL	Puting Beliung	0	0	0	0	0	0	0	4	0	0	0	0
10	3517	JOMBANG	PUTING BEL	Puting Beliung	0	1	0	0	5	5	0	8	0	0	0	0
11	3512	SITUBONDO	PUTING BEL	Puting Beliung	0	1	0	0	0	1	0	0	0	0	0	0
12	3507	MALANG	PUTING BEL	Puting Beliung	0	0	0	0	0	0	2	28	0	0	0	0
13	3515	SIDOARJO	PUTING BEL	puting beliung	0	0	0	0	0	0	0	11	0	0	0	0
14	3518	NGANJUK	PUTING BEL	puting beliung	0	0	0	0	0	0	0	18	0	0	0	0
15	3507	MALANG	PUTING BEL	Puting Beliung	0	0	0	0	0	0	0	12	1	0	0	0
16	3513	PROBOLINGGO	PUTING BEL	Puting Beliung	0	0	0	0	0	1	0	0	0	0	0	0
17	3515	SIDOARJO	PUTING BEL	Angin puting	0	0	0	0	0	0	24	0	0	0	0	0
18	3519	MADIUN	PUTING BEL	angin kencang	0	0	0	0	0	0	0	9	0	0	0	0
19	3502	PONOROGO	PUTING BEL	Angin Puting	0	0	0	0	0	0	0	2	0	0	0	0
20	3514	PASURUAN	PUTING BEL	Puting beliung	0	0	0	0	0	0	0	0	0	0	0	0
21	3508	LUMAJANG	PUTING BEL	Puting beliung	0	0	0	0	0	0	1	0	0	0	0	0
22	3512	SITUBONDO	PUTING BEL	puting beliung	0	0	0	4	0	0	1	0	0	0	0	0
23	3513	PROBOLINGGO	PUTING BEL	puting beliung	0	0	0	0	0	2	6	68	0	0	0	0
24	3508	LUMAJANG	PUTING BEL	puting beliung	0	0	0	3	0	0	1	0	0	0	0	0
25	3529	SUMENEP	PUTING BEL	puting beliung	0	0	0	0	0	0	0	3	0	0	0	0
26	3515	SIDOARJO	PUTING BEL	puting beliung	0	0	0	0	0	0	1	0	0	0	0	0

Gambar 3.1 Data Bencana Alam di Jawa Timur

(Sumber : <http://dibi.bnpb.go.id/>)

3.2 Struktur Data dan Variabel Penelitian

Pada dasarnya, penelitian ini menggunakan dua macam data yaitu:

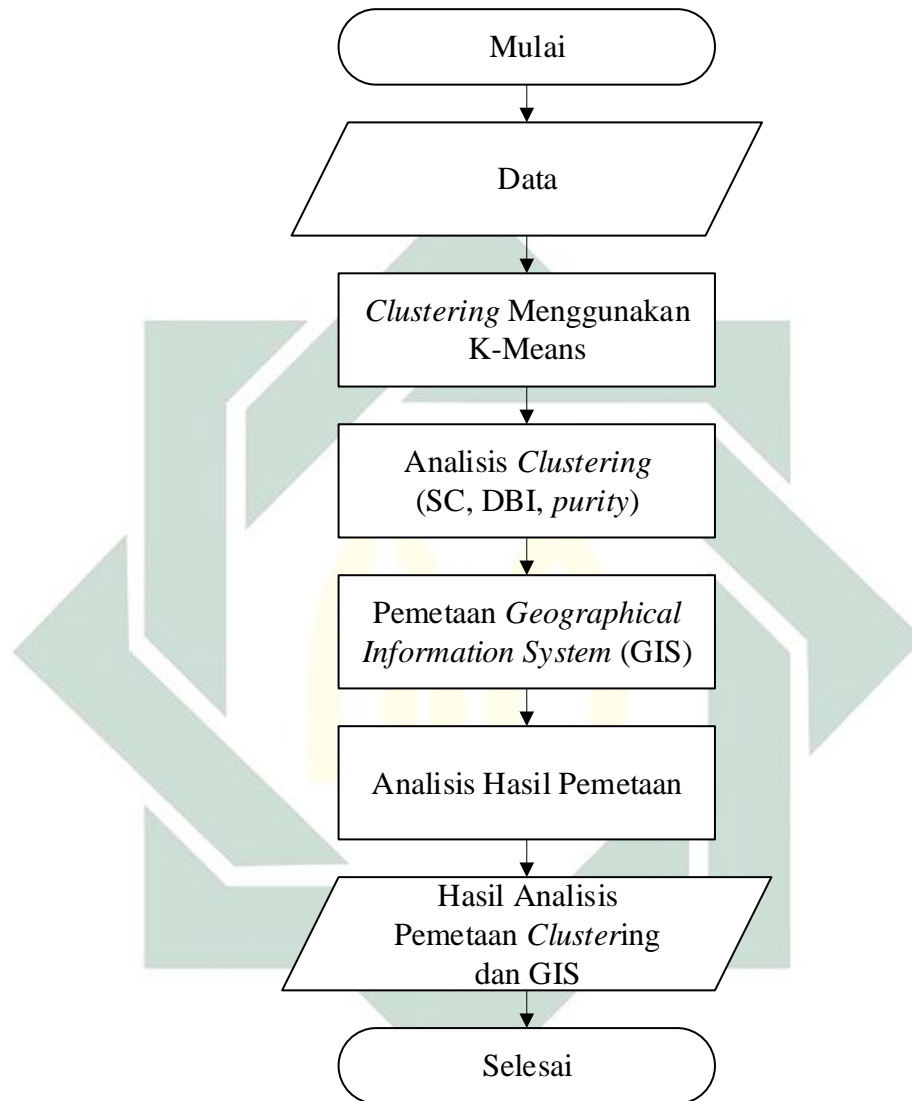
1. Data Spasial

Data spasial adalah gambaran nyata atau peta suatu wilayah yang terdapat di permukaan bumi. Umumnya direpresentasikan berupa grafik, peta, gambar dengan format digital dan disimpan dalam bentuk koordinat x,y (*vector*) atau dalam bentuk *image* (raster) yang memiliki nilai tertentu.

2. Data Non Spasial (Atribut) yaitu data berbentuk tabel dimana tabel tersebut berisi informasi-informasi yang dimiliki oleh obyek dalam data spasial. Data tersebut berbentuk data tabular yang saling terintegrasi dengan data spasial yang ada. Contohnya ialah korban yang terdampak akibat bencana alam, kerusakan yang dialami setelah bencana alam, dan lain-lain. Berikut merupakan variabel-variabel yang akan digunakan pada penelitian ini:

Tabel 3.1 Definisi Operasional Variabel

No.	Variabel	Definisi	Satuan
1	Jumlah Kejadian Puting Beliung	Jumlah kejadian bencana puting beliung	Jumlah Kejadian
2	Korban Bencana	Jiwa	Jumlah Korban
3	Kerusakan Rumah	Kerusakan bangunan roboh	Jumlah Rumah
4	Kerusakan Unit (Fasilitas)	Kerusakan Unit harta benda	Jumlah unit yang rusak



Gambar 3.2 Diagram Alir Penelitian

- b. Setelah data diperoleh, maka dilakukan *clustering* menggunakan algoritma K-Means. Dimana dapat diketahui kelompok-kelompok dari setiap daerah manakah yang rawan akan bencana alam.
- c. Setelah terbentuk *cluster*, langkah selanjutnya yaitu melakukan uji validasi *cluster*. Hal ini dilakukan untuk mengetahui kekuatan struktur data dari jumlah *cluster*.

Pada penelitian akan menggunakan beberapa metode evaluasi *cluster* diantaranya yaitu *silhouette coefficient*, *davies bouldin index*, dan *purity*.

- d. Langkah terakhir yaitu dilakukan pewarnaan, dalam hal ini pewarnaan dilakukan dengan GIS berdasarkan peta Jawa Timur. Pada GIS ini akan dilakukan pewarnaan sesuai dengan hasil kerawanan bencana puting beliung.
- e. Setelah pewarnaan, dilakukan analisis hasil pewarnaan berdasarkan zonasi GIS. Hasil akhir dari penelitian adalah analisis karakteristik dengan zonasi wilayah rawan bencana angin puting beliung di Jawa Timur menggunakan algoritma K-Means dengan analisis *silhouette coefficient*, *davies bouldin index*, dan *purity*.

Tabel 4.3 Hasil Perhitungan Jarak Euclidean Iterasi 1

No.	Kab/Kota	C1	C2	C3	C4	Hasil
1	Banyuwangi	1256.84	0	3	526.91	C2
2	Bojonegoro	3290.36	244.10	141.23	366.23	C3
3	Jombang	0	1446.10	2874.58	471.37	C1
4	Lamongan	1365	3	0	546.54	C3
5	Madiun	1325.68	0	3	548.02	C2
6	Malang	1446.10	0	3	525.55	C2
7	Pacitan	1145.70	0	3	538.27	C2
8	Probolinggo	2874.58	3	0	274.52	C3
9	Sumenep	980.96	0	3	528.38	C2
10	Tulung Agung	5654.45	525.55	274.52	0	C4

Berdasarkan Tabel 4.3 dapat diketahui bahwa data ke-1 (Kab. Banyuwangi) masuk ke dalam *cluster* 2, karena data ke-1 memiliki jarak minimum atau terdekat dengan *centroid* 2. Hal ini diterapkan untuk semua data. Langkah selanjutnya yaitu melakukan iterasi ke-2 dengan menentukan *centroid* baru. *Centroid* terbaru ditentukan dengan menghitung rata-rata setiap *cluster*, untuk menghitung maka dapat dilihat pada perhitungan berikut ini:

Centroid 1 terdiri dari anggota pada *cluster* 1 yaitu Kabupaten Jombang.

Tabel 4.4 *Centroid Cluster* 1

	a	b	c	d	e	f	g	h	i	j	k
C1	0	1	0	14	0	0	1	118	0	13	3

Pada Tabel 4.7 terdapat 1 data yang mengalami perpindahan *cluster* yaitu Kabupaten Lamongan yang terletak pada data ke-4. Maka dilakukan iterasi selanjutnya hingga data tidak berpindah ke *cluster* lain yaitu dengan melakukan rata-rata setiap *cluster* untuk menentukan *centroid* selanjutnya.

Tabel 4.8 Centroid Terbaru

Cluster Ke-	a	b	c	d	e	f	g	h	i	j	k
C1	0	1	0	14	0	0	1	118	0	13	3
C2	0.17	1.17	0	10.83	0	8.17	4.83	24.5	0.33	0	0.5
C3	1	0.5	0	252.5	0	9.5	10	118	0	0.5	1.5
C4	0	0	0	390	0	0	0	402	0	1	0

Setelah menemukan *centroid* seperti pada Tabel 4.8 maka dilakukan iterasi ke-3 dan menghasilkan *cluster* seperti pada Tabel 4.9 berikut.

Tabel 4.9 Iterasi ke-3

No.	Kab/Kota	C1	C2	C3	C4	Hasil	Iterasi Ke-2
1	Banyuwangi	97.53	15.8	247.18	526.91	C2	C2
2	Bojonegoro	274.45	270.74	70.62	366.22	C3	C3
3	Jombang	0	94.91	239.19	471.36	C1	C1
4	Lamongan	105.80	31.3	271.90	546.53	C2	C2
5	Madiun	102.79	16.27	272.30	548.02	C2	C2
6	Malang	111.98	33.19	238.61	525.55	C2	C2
7	Pacitan	89.06	15.83	267.43	538.28	C2	C2
8	Probolinggo	221.5	266.50	70.62	274.52	C3	C3
9	Sumenep	76.55	27.62	262.73	528.38	C2	C2
10	Tulung Agung	471.35	535.13	315.84	0	C4	C4

Tabel 4.9 menunjukkan hasil *cluster* tidak berubah atau sama dengan hasil *cluster* iterasi sebelumnya. Artinya, iterasi berakhir karena data tidak berpindah ke *cluster* lain. Maka dapat disimpulkan bahwa anggota *cluster* 1 yaitu Jombang, *cluster* 2 yaitu Banyuwangi, Lamongan, Madiun, Malang, Pacitan, dan Sumenep. *Cluster* 3 yaitu Bojonegoro dan Probolinggo. Untuk *cluster* 4 yaitu Tulung Agung. *Cluster* yang

telah terbentuk memiliki karakteristik masing-masing. Untuk mengetahui karakteristik setiap *cluster* dapat di ilustrasikan setiap *cluster* menggunakan data awal *clustering*.

Tabel 4.10 Hasil *Cluster 1*

		C1										
Kab/Kota	Data Ke-	a	b	c	d	e	f	g	h	i	j	k
Jombang	1	0	1	0	14	0	0	1	118	0	13	3

Cluster 1 hanya beranggotakan 1 Kabupaten saja yaitu Kabupaten Jombang, dimana daerah tersebut tergolong daerah aman karena bencana yang terjadi tidak berdampak cukup berat seperti daerah lain. Dengan banyak data pada Rumah Rusak Ringan (RRR) sebanyak 118.

Tabel 4.11 Hasil *Cluster 2*

		C2										
Kab/Kota	Data Ke-	a	b	c	d	e	f	g	h	i	j	k
Banyuwangi	1	0	0	0	25	0	3	1	22	0	0	0
Lamongan	2	0	0	0	0	0	37	8	21	1	0	3
Madiun	3	0	0	0	0	0	0	0	17	0	0	0
Malang	4	1	5	0	40	0	4	2	10	0	0	0
Pacitan	5	0	1	0	0	0	0	0	31	1	0	0
Sumenep	6	0	1	0	0	0	5	18	46	0	0	0

Cluster 2 beranggotakan 6 daerah yaitu Banyuwangi, Lamongan, Madiun, Malang, Pacitan dan Sumenep. Dengan kerusakan dan korban yang cukup banyak maka daerah pada *cluster 2* tergolong daerah “cukup rawan”.

Tabel 4.12 Hasil *Cluster 3*

		C3										
Kab/Kota	Data Ke-	a	b	c	d	e	f	g	h	i	j	k
Bojonegoro	1	0	0	0	280	0	11	10	53	0	1	0
Sumenep	2	0	1	0	0	0	5	18	46	0	0	0

Cluster 3 beranggotakan 2 daerah yaitu Bojonegoro dan Sumenep, dimana daerah ini memiliki karakteristik “rawan” karena data terdampak cukup banyak yaitu 280.

Tabel 4.13 Hasil *Cluster* 4

		C4										
Kab/Kota	Data Ke-	a	b	c	d	e	f	g	h	i	j	k
Tulung Agung	1	0	0	0	390	0	0	0	402	0	1	0

Cluster 4 memiliki anggota 1 daerah yang tergolong sangat rawan karena sesuai kejadian yang pernah terjadi, tercatat terdapat korban terdampak 390 dan rumah rusak ringan sebanyak 402.

4.2 Contoh Perhitungan Manual Evaluasi *Clustering*

Evaluasi *cluster* dilakukan dengan tujuan untuk mengetahui seberapa kuat struktur data pada *cluster* yang telah terbentuk. Penelitian ini menggunakan 3 metode evaluasi *cluster* diantaranya *silhouette coefficient*, *davies bouldin index* dan *purity*.

4.2.1 *Silhouette Coefficient*

Untuk memulai perhitungan manual evaluasi *cluster* menggunakan *silhouette coefficient*, dibutuhkan data yang telah terbentuk dalam *clustering* yang telah dilakukan. Hal ini dapat dilihat pada Tabel 4.10 hingga 4.13. Seperti yang telah disebutkan pada bab II, untuk mencari nilai *silhouette coefficient* terdapat beberapa langkah yang telah dijelaskan pada Persamaan 2.4, 2.5, 2.6, dan 2.7. Tahap pertama yaitu mencari nilai $a(i)$ menggunakan Persamaan 2.4, nilai $a(i)$ diperoleh dari perhitungan rata-rata dari perhitungan jarak data ke data yang lain dalam satu *cluster*. Berikut perhitungan dari data pada *cluster* 1 yaitu Kabupaten Jombang.

$$\begin{aligned}
 a(i) &= \frac{1}{|10|} \sqrt{(0-0)^2 + (1-0)^2 + (0-0)^2 + (14-0)^2 + (0-0)^2 +} \\
 &\quad \sqrt{(0-0)^2 + (1-0)^2 + (118-0)^2 + (0-0)^2 +} \\
 &\quad \sqrt{(13-0)^2 + (3-0)^2} \\
 &= 118.8192 \\
 &= 118.82
 \end{aligned}$$

Langkah kedua yaitu mencari nilai $d(i,C)$. Nilai $d(i,C)$ merupakan perhitungan jarak data Kab. Jombang (*Cluster 1*) dengan data pada *cluster* lain. Perhitungan pertama yaitu antara Jombang (*Cluster 1*) dan Banyuwangi (*Cluster 2*) seperti perhitungan berikut ini.

$$\begin{aligned}
 d(i,C) &= \frac{1}{|6|} \sqrt{(0-0)^2 + (1-0)^2 + (0-0)^2 + (14-25)^2 + (0-0)^2 +} \\
 &\quad \sqrt{(0-3)^2 + (1-1)^2 + (118-22)^2 + (0-0)^2 +} \\
 &\quad \sqrt{(13-0)^2 + \dots (i-j)^2} \\
 &= 109.83
 \end{aligned}$$

Perhitungan $d(i,C)$ akan dilanjutkan untuk seluruh data pada *cluster* lain. Untuk langkah selanjutnya yaitu menghitung nilai $b(i)$ seperti berikut ini.

$$\begin{aligned}
 b(i) &= \min(109.83, 163.92, \dots) \\
 &= 109.83
 \end{aligned}$$

Langkah setelah mencari nilai minimum $d(i,C)$ yaitu menghitung nilai $s(i)$ atau *silhouette*.

$$\begin{aligned}
 s(1) &= \frac{(109.83 - 118.82)}{\max(109.83 - 118.82)} \\
 &= 0.99
 \end{aligned}$$

Hal ini dilakukan sampai $s(10)$ dan menghasilkan nilai seperti Tabel 4.14.

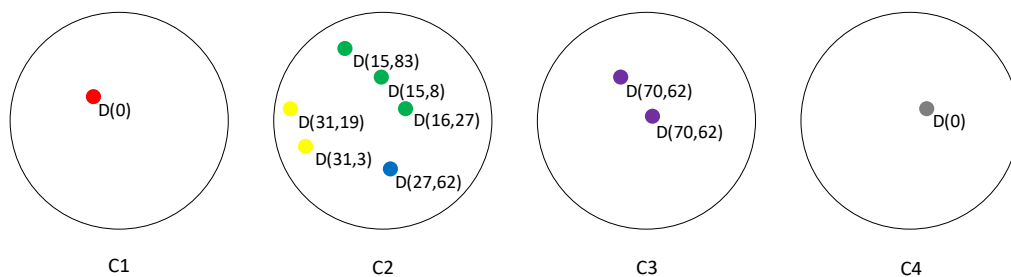
Tabel 4.14 Hasil Contoh Perhitungan Nilai *Silhouette Coefficient*

Cluster ke -	Kab/Kota	a(i)	b(i)	s(i)
1	Jombang	118.82	109.83	0.99
	Banyuwangi	19.26	29.19	0.63
	Lamongan	37.84	29.19	0.90
2	Madiun	34.51	100.52	0.66
	Malang	24.91	107.30	0.77
	Pacitan	21.31	86.27	0.75
	Sumenep	33.73	23.98	-0.29
3	Probolinggo	129.88	68.95	-0.47
	Bojonegoro	130.30	80.27	-0.38
4	Tulungagung	402.49	284.63	-0.29

Hasil akhir dari *silhouette coefficient* adalah dengan menghitung rata-rata nilai $s(i)$ seluruh data yaitu 0.327.

4.2.2 Purity

Untuk mengetahui nilai *purity* maka diperlukan hasil *clustering* akhir dan jarak *euclidean* pada iterasi terakhir *clustering*, seperti yang telah ditunjukkan pada Tabel 4.9. Gambar 4.1 merupakan ilustrasi hasil *clustering* yang telah terbentuk menggunakan K-Means *clustering*.



Gambar 4.1 Ilustrasi Hasil *Clustering*

Gambar 4.1 merupakan ilustrasi hasil *clustering* dimana akan dihitung nilai *purity* menggunakan Persamaan 2.13.

$$\begin{aligned} Purity &= \frac{1}{10} (1 + 3 + 2 + 1) \\ &= 0.7 \end{aligned}$$

Seperti yang telah dijelaskan pada bab II bahwa *cluster* dikatakan optimal jika nilai *purity* mendekati 1 atau sama dengan 1. Perhitungan ini akan dilakukan melalui beberapa kali percobaan untuk data yang cukup banyak, maka dari itu diperlukan alat pengolahan data yang tepat untuk mempercepat penyelesaian perhitungan.

4.2.3 Davies Bouldin Index (DBI)

Untuk mendapatkan nilai *davies bouldin index*, terlebih dahulu dihitung nilai *sum of square within-cluster*, *sum of square between-cluster* dan *ratio*. Untuk memulai perhitungan manual *davies bouldin index* diperlukan data yang sudah terbentuk pada *clustering* akhir dan titik pusat (*centroid*) *cluster* yang terakhir seperti pada Tabel 4.9. Setelah mengetahui jarak setiap data dengan *centroid cluster* yang didapat dari iterasi terakhir pada *clustering*, selanjutnya akan dilakukan perhitungan nilai *Sum of Square Within Cluster (SSW)* dengan Persamaan 2.8.

$$SSW_1 = 0$$

$$SSW_2 = \frac{15.8+31.3+16.27+33.19+15.83+27.62}{6} = 23.34$$

$$SSW_3 = \frac{70.62}{1} = 70.62$$

$$SSW_4 = 0$$

Setelah mengetahui nilai SSW maka dilakukan perhitungan *Sum of Square Between Cluster* (SSB). Untuk menghitung nilai SSB maka diperlukan *centroid* terakhir pada iterasi terakhir. Berikut Tabel 4.15 *centroid* akhir yang di dapat dari iterasi terakhir.

Tabel 4.15 Titik pusat *cluster* akhir dari proses *clustering*

Titik Pusat	a	b	c	d	e	f	g	h	i	j	k
C1	0	1	0	14	0	0	1	118	0	13	3
C2	0.17	1.17	0	10.83	0	8.17	4.83	24.5	0.33	0	0.5
C3	1	0.5	0	252.5	0	9.5	10	118	0	0.5	1.5
C4	0	0	0	390	0	0	0	402	0	1	0

Untuk menghitung nilai *Sum of Square Between Cluster* (SSB) menggunakan Persamaan 2.9.

$$\begin{aligned}
 SSB_{1,2} &= \sqrt{(0 - 0.17)^2 + (1 - 1.17)^2 + (0 - 0)^2 + (14 - 10.83)^2 + (0 - 0)^2 +} \\
 &\quad \sqrt{(0 - 8.17)^2 + (1 - 4.83)^2 + (118 - 24.5)^2 + (0 - 0.33)^2 + (13 - 0)^2 +} \\
 &\quad \sqrt{(3 - 0.5)^2} \\
 &= 94.91645 \\
 &= 94.92
 \end{aligned}$$

Perhitungan ini dilakukan hingga seluruh *centroid* terpenuhi.

Tabel 4.16 Nilai *Sum of Square Between Cluster*

SSB _{1,2}	94.91645
SSB _{1,3}	239.1929
SSB _{1,4}	471.3672
SSB _{2,3}	259.1865
SSB _{2,4}	535.1345
SSB _{3,4}	315.8421

Setelah dihitung nilai SSW dan SSB, akan dilakukan perhitungan rasio antar *cluster* ke-*i* dan *cluster* ke-*j* ($R_{i,j}$) menggunakan Persamaan 2.10.

$$R_{1,2,3,4} = \frac{SSW_1 + SSW_2 + SSW_3 + SSW_4}{SSB_{1,2} + \dots + SSB_{3,4}}$$

$$R_{1,2,3,4} = \frac{0 + 23.34 + 70.62 + 0}{94.92 + 239.2 + 471.4 + 259.18 + 535.13 + 315.84}$$

$$= 0.05$$

Langkah terakhir dari metode *davies bouldin index* yaitu menentukan rasio antar *cluster* maksimum R(max). Setelah mengetahui semua rasio antar *cluster* maksimum seperti Tabel 4.17.

Tabel 4.17 Nilai Rasio Maksimum

Rasio	12	13	23	24	34	Rmax
2 Cluster	0.25	0.29	0.18	-0.04	-0.22	0.29
Rasio	123	124	234	134		
3 Cluster	0.16	0.02	0.07	0.08		0.16
Rasio	1234					
4 Cluster	0.05					0.05

Setelah diketahui nilai rasio antar *cluster* maksimum, selanjutnya dihitung nilai DBI. Untuk menghitung nilai DBI menggunakan Persamaan 2.11.

$$DBI = \frac{R_1 + R_2 + R_3}{k}$$

$$= \frac{0.29 + 0.16 + 0.05}{4}$$

$$= 0.12$$

Setelah melakukan analisis perhitungan evaluasi *cluster* menggunakan DBI dengan percobaan 4 *cluster* di dapatkan nilai DBI sebesar 0.20. Selanjutnya perhitungan akan dilakukan menggunakan program. Karena jumlah data yang cukup banyak dan untuk mempercepat penyelesaian.

4.3 Hasil dan Analisis *Clustering* Menggunakan Metode K-Means

Hasil pengelompokan kota dan kabupaten rawan bencana angin puting beliung di Jawa Timur selama tahun 2019 sebanyak 4 *cluster* menggunakan metode K-Means ditunjukkan Tabel 4.18.

Tabel 4. 18 Hasil *Cluster* Kota/Kab. di Jawa Timur Tahun 2019

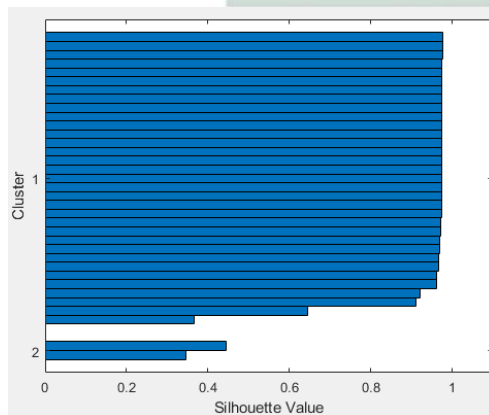
Kelompok	Jumlah Anggota	Kota/Kabupaten
1	33	Bangkalan, Lamongan, Banyuwangi, Batu, Blitar, Bondowoso, Gresik, Jember, Jombang, Kediri, Pasuruan, Kota Batu, Kota Blitar, Kota Madiun, Kota Malang, Lumajang, Madiun, Magetan, Surabaya, Malang, Mojokerto, Ngawi, Pacitan, Pamekasan, Ponorogo, Situbondo, Sumenep, Trenggalek, Tuban, Sampang
2	1	Tulung Agung
3	3	Probolinggo, Bojonegoro, Nganjuk
4	1	Sidoarjo

Setelah dilakukan *clustering* menggunakan metode K-Means, maka dilakukan uji validasi atau evaluasi *cluster*. Untuk menentukan *cluster* yang tepat maka perlu dilakukan percobaan atau uji validasi jumlah *cluster* yang tepat seperti pada Tabel 4.19.

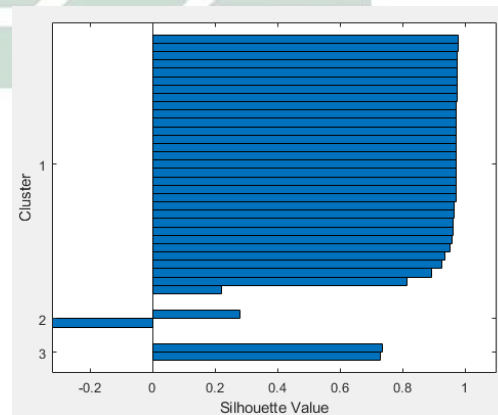
Tabel 4.19 Tabel Uji Validasi *cluster*

Jumlah Cluster	Silhouette	Davies Bouldin Index	Purity
2	0.8678	0.6377	0.8857
3	0.9052	0.4799	0.9714
4	0.9116	0.3633	1
5	0.718	0.4758	0.9429
6	0.7149	0.4555	0.9714

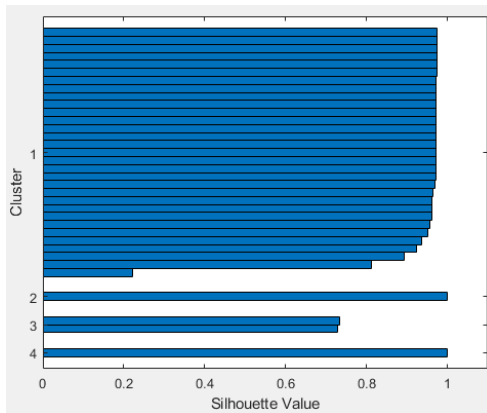
Tabel 4.19 merupakan uji evaluasi *cluster* menggunakan 3 metode yaitu *silhouette coefficient*, *davies bouldin index*, dan *purity*. Nilai pada Tabel 4.19 menunjukkan bahwa ketiga metode tersebut menunjukkan hasil *cluster* yang optimal terletak pada jumlah *cluster* sebanyak empat. Nilai *silhouette* yang mendekati 1 atau sama dengan 1 disebut struktur data yang kuat atau jumlah *cluster* tersebut optimal. Pada Tabel 4.19 nilai *silhouette* terbesar adalah 0.9116 yang terletak pada jumlah *cluster* sebanyak empat. Grafik *silhouette* dari nilai $s(i)$ setiap data di representasikan seperti pada Gambar 4.3-4.8.



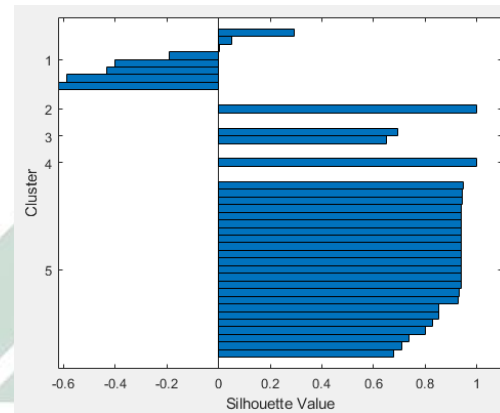
Gambar 4.2 K-Means dengan 2 Cluster



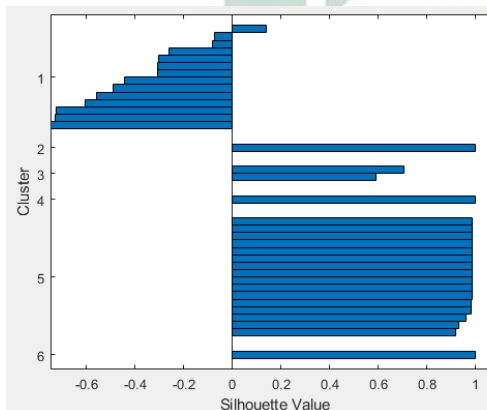
Gambar 4.3 K-Means dengan 3 Cluster



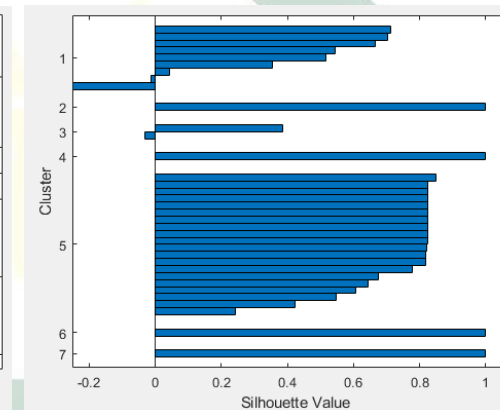
Gambar 4.4 K-Means dengan 4 Cluster



Gambar 4.5 K-Means dengan 5 Cluster



Gambar 4.6 K-Means dengan 6 Cluster



Gambar 4.7 K-Means dengan 7 Cluster

Gambar 4.3 hingga 4.8 menunjukkan keakuratan jumlah setiap *cluster*, dimana *cluster* tersebut didapatkan dari nilai *silhouette coefficient* dan rata-ratanya, sehingga diketahui manakah *cluster* terbaik untuk struktur data yang kuat. Untuk menentukan *cluster* terbaik menggunakan *silhouette* dapat ditinjau dari berbagai aspek yaitu grafik *silhouette coefficient* pada Gambar 4.3 hingga 4.8. Jika pada grafik terdapat hasil *cluster* yang berada disebelah kiri 0 artinya nilai *silhouette* kurang dari 0 atau minus. Hal tersebut menandakan data *overlapping* atau data tidak memiliki *cluster* yang tepat. Jadi tidak dapat disimpulkan bila rata-ratanya memiliki nilai yang tinggi maka *cluster*

tersebut baik. Hal ini harus di pertimbangkan dengan melihat grafik *silhouette* pada Gambar 4.3 hingga 4.8. Oleh karena itu, dapat disimpulkan bahwa *cluster* terbaik adalah 4 *cluster*. Karena dilihat dari Gambar 4.6 hampir seluruh data terletak pada nilai 1, artinya struktur data tersebut sangat kuat dan jika ditinjau dari reratanya maka *cluster* yng berjumlah empat.

Untuk validasi *cluster* menggunakan *purity*, *cluster* dikatakan optimal jika bernilai 1 atau mendekati 1. *Purity* memiliki nilai kemiripan dengan *silhouette*, semakin tinggi nilainya mengartikan semakin optimal jumlah *cluster* tersebut. Percobaan evaluasi *cluster* menggunakan *purity* ini dilakukan sebanyak 5 kali, dengan jumlah *cluster* sebanyak 2 hingga 6. Berikut hasil perhitungan *purity* menggunakan program.

Dari Tabel 4.19 dapat diketahui bahwa hasil *cluster* optimal yaitu 4 *cluster*. Dimana semakin mendekati 1 atau sama dengan 1 artinya *cluster* tersebut murni dan memiliki kedekatan antar data yang cukup baik.

Untuk validasi *cluster* menggunakan *davies bouldin index*, *cluster* dikatakan optimal jika nilai DBI tersebut mendekati 0. Dapat dilihat pada Tabel 4.19 bahwa *cluster* yang optimal adalah *cluster* 4. Karena jika dilihat nilai DBI pada *cluster* 2 sampai 6, nilai yang mendekati 0 adalah *cluster* 4 dengan nilai 0.3633. Pada dasarnya, nilai DBI didasarkan oleh nilai *inter-cluster* dan *intra-cluster*. Dimana untuk mendapatkan nilai tersebut melalui perhitungan jarak kedekatan data terhadap *cluster* yang terbentuk dengan *cluster* yang lain atau nilai *sum of square within cluster* dan *sum of square between cluster*. Semakin kecil nilai *inter-cluster* dan *intra-cluster* berarti

masing-masing objek dalam *cluster* tersebut memiliki tingkat kemiripan dan kesamaan karakteristik yang tinggi.

4.4 Interpretasi *Cluster* Optimal

Langkah awal *clustering* menggunakan K-Means yaitu dengan melakukan pemilihan *centroid* secara acak. Setelah iterasi pertama akan dilanjutkan dengan pemilihan *centroid* menggunakan *mean* atau rata-rata. Dengan adanya *centroid* yang telah terpilih maka dapat dilihat pola pembentukan masing-masing *cluster*. Berikut *centroid* akhir pada *clustering* K-Means menggunakan data terdampak bencana angin puting beliung tahun 2019 di Jawa Timur. *Centroid* akhir digunakan untuk mengetahui kedekatan antar *cluster*. Jika menghasilkan jarak *euclidean* sangat minimum artinya *cluster* yang terbentuk pada *cluster* tersebut memiliki kedekatan antar data yang cukup baik.

Tabel 4.20 *Centroid* Akhir Pembentukan *Cluster* Menggunakan K-Means

<i>Cluster</i>	a	b	c	d	e	f	g	h	i	j	k
C1	0.10	0.47	0.00	7.43	0.37	3.90	3.00	23.60	0.10	0.73	0.43
C2	0.00	0.00	0.00	390.00	0.00	0.00	0.00	402.00	0.00	1.00	0.00
C3	0.67	0.33	0.00	216.67	0.00	6.67	9.33	93.33	0.33	0.33	1.33
C4	0.00	0.00	0.00	12.00	0.00	0.00	26.00	507.00	0.00	0.00	0.00

Tabel 4.20 merupakan *centroid* akhir pada pembentukan *cluster* dimana dapat dianalisa bahwa setiap *centroid* memiliki pola tertentu. Hal ini akan direpresentasikan berdasarkan pola nilai *centroid* seperti pada Tabel 4.21.

Tabel 4. 21 Pola Pada *Centroid* Akhir

<i>Cluster</i>	a	b	d	e	f	g	h	i	j	k
C1	sedang	tinggi	rendah	tinggi	sedang	sedang	rendah	sedang	tinggi	sedang
C2	rendah	rendah	sangat tinggi	rendah	rendah	rendah	tinggi	rendah	sangat tinggi	rendah
C3	tinggi	sedang	tinggi	rendah	tinggi	tinggi	sedang	tinggi	sedang	tinggi
C4	rendah	rendah	sedang	rendah	rendah	sangat tinggi	sangat tinggi	rendah	rendah	rendah

Pada variabel a (korban meninggal) dengan *range* data 0-2, *cluster* 1 terkategoriikan sedang dengan nilai 0.1, *cluster* 2 dan 4 termasuk kategori rendah dengan nilai 0 dan *cluster* 4 termasuk kategori tinggi dengan nilai 0.67. Variabel b (korban luka-luka) memiliki *range* data 0-2, dimana *cluster* 1 termasuk kategori tinggi dengan nilai 0.47, *cluster* 2 dan 4 termasuk kategori rendah dengan nilai 0 dan *cluster* 3 termasuk kategori sedang. Variabel c (korban hilang) memiliki nilai yang sama maka tidak berpengaruh terhadap terbentuknya *cluster*.

Pada variabel d (korban terdampak) dengan *range* 0-6, *cluster* 1 termasuk dalam kategori tinggi dengan nilai 0.37 dan *cluster* 2,3, dan 4 termasuk dalam kategori rendah dengan nilai 0. Variabel f (rumah rusak berat) dengan *range* data 0-37 menunjukkan bahwa *cluster* 1 memiliki kategori sedang, *cluster* 2 dan 4 termasuk dalam kategori rendah dengan nilai 0, *cluster* 3 termasuk dalam kategori tinggi dengan nilai 6.67. Variabel g (rumah rusak sedang) memiliki *range* data 0-26 dimana *cluster* 1 memiliki kategori sedang dengan nilai 3, *cluster* 2 memiliki kategori rendah dengan

nilai 0, *cluster* 3 termasuk dalam kategori tinggi dengan nilai 9.33 dan *cluster* 4 termasuk dalam kategori sangat tinggi dengan nilai 26.

Variabel h (rumah rusak ringan) memiliki *range* data 0-507 dimana *cluster* 1 termasuk dalam kategori rendah dengan nilai 23.6, *cluster* 2 dengan nilai 402 terkategori tinggi, *cluster* 3 memiliki kategori sedang dengan nilai 93.33, dan *cluster* 4 memiliki kategori sangat tinggi dengan nilai 507. Untuk variabel j (fasilitas peribadatan) dengan *range* data 0-13, *cluster* 1 memiliki nilai 0.73 dengan kategori tinggi, *cluster* 2 memiliki nilai 1 dengan kategori sangat tinggi, kategori sedang terdapat pada *cluster* 3 dengan nilai 0.33 dan kategori rendah terdapat pada *cluster* 4 dengan nilai 0. Untuk variabel k (fasilitas pendidikan) dengan *range* data 0-3 kategori rendah terdapat pada *cluster* 2 dan 4 dengan nilai 0, kategori sedang terdapat pada *cluster* 1 dengan nilai 0.43, dan untuk *cluster* 3 termasuk kedalam kategori tinggi dengan nilai 1.33.

Untuk mengetahui karakteristik terbentuknya *cluster* kerawanan bencana angin puting beliung di Jawa Timur maka dilakukan interpretasi berdasarkan data bencana angin puting beliung di Jawa Timur tahun 2019. Ditinjau dari korban bencana yang terjadi, hasil *cluster* dan data yang telah diolah maka dapat diketahui bahwa daerah-daerah yang jarang terjadi bencana terletak pada *cluster* 1 dengan anggota berjumlah 33 kabupaten/kota. Pada *cluster* 2 terdapat 1 anggota dimana jika ditinjau dari data tersebut, 1 kabupaten tersebut memiliki nilai yang paling tinggi pada korban bencana terdampak yaitu 390. Hal ini artinya Kabupaten Tulung Agung termasuk dalam zona

sangat rawan bencana angin puting beliung. Untuk *cluster* 3 beranggotakan 3 yaitu Kota/Kab. Probolinggo, Nganjuk dan Bojonegoro. Jika ditinjau dari data pada variabel korban, nilai dari 3 daerah tersebut memiliki nilai cukup dibandingkan yang lain yaitu range 145-280 artinya daerah ini cukup rawan akan bencana angin puting beliung. Sedangkan *cluster* 4 hanya berjumlah satu anggota yaitu Sidoarjo.

Ditinjau dari kerusakan rumah, maka dapat dianalisis bahwa *cluster* 1 merupakan daerah aman, *cluster* 2 merupakan daerah rawan karena terdapat rumah rusak dengan kategori kerusakan ringan berjumlah 402 rumah. Untuk *cluster* 3 memiliki karakteristik cukup rawan dengan 3 anggota yaitu Probolinggo, Nganjuk, Bojonegoro. Untuk *cluster* 4 tergolong daerah sangat rawan karena terdapat kerusakan rumah dengan kategori sedang berjumlah 26 rumah dan kategori ringan sebesar 507, ini merupakan nilai tertinggi dibandingkan daerah lain.

Ditinjau dari kerusakan fasilitas umum, terlihat dari data dan *cluster* yang terbentuk bahwa *cluster* 1 menunjukkan terdapat nilai yang cukup tinggi yaitu pada Kabupaten Jombang dengan kerusakan peribadatan sebanyak 12 tempat dan pendidikan sebanyak 3 tempat. Maka dari itu *cluster* 1 tergolong daerah sangat rawan bencana untuk kategori kerusakan fasilitas umum. Untuk *cluster* 2 tergolong daerah yang cukup rawan dengan nilai 1 tempat untuk kerusakan peribadatan rumah. *Cluster* 3 memiliki karakteristik daerah rawan dengan nilai Bojonegoro sebanyak 1 tempat kerusakan yaitu tempat peribadatan. Nganjuk dengan nilai kerusakan sebanyak 1 tempat kesehatan dan 1 tempat pendidikan. Probolinggo dengan nilai 3 tempat

pendidikan rusak. Untuk *cluster* 4 memiliki karakteristik aman untuk kategori kerusakan fasilitas umum. Yaitu Kabupaten Sidoarjo dengan tidak ada kerusakan fasilitas umum.

4.5 Pewarnaan dan Analisis Zonasi Angin Puting Beliung

Langkah akhir dari penelitian ini adalah pewarnaan atau zonasi daerah rawan bencana angin puting beliung Jawa Timur. Setelah dilakukan *clustering* dan validasinya maka akan dilakukan pewarnaan menggunakan *Geographical Information System* (GIS). Dalam hal ini, setelah diketahui manakah kabupaten dan kota yang tergolong dalam kelompok rawan, aman maupun parah akan dianalisa menggunakan data peta geografis maupun demografi yang bersumber dari beberapa dinas terkait. Gambar 4.8 merupakan hasil pewarnaan daerah rawan puting beliung menggunakan data terdampak tahun 2019 di Jawa Timur.

Berdasarkan Gambar 4.8 dapat diketahui bahwa Kabupaten Tulung Agung merupakan daerah dengan kondisi sangat rawan. Kondisi tersebut ditinjau dari data terdampak, dimana Tulung Agung merupakan daerah yang memiliki nilai korban terdampak terbesar yaitu 390 korban terdampak dan 402 rumah rusak ringan. Hal ini dipengaruhi oleh beberapa faktor. Yuniarti (2018) menyebutkan faktor yang signifikan berpengaruh terhadap kejadian puting beliung yaitu variabel suhu minimum H-1 kejadian, curah hujan rata-rata H-1 kejadian, kelembapan udara H-1 kejadian, suhu minimum hari-H kejadian dan curah hujan rata-rata hari H-kejadian.

Pada Gambar 4.8 Probolinggo, Bojonegoro dan Nganjuk merupakan daerah yang tergolong cukup rawan karena dilihat dari data terdampak 2019 daerah tersebut memiliki kerusakan yang cukup banyak yaitu pada korban terdampak dan rumah rusak ringan. BPBD Probolinggo mengatakan cuaca ekstrim sangat berpengaruh akan terjadinya angin puting beliung. Maka dari itu, masyarakat diharapkan lebih waspada dengan hal tersebut. Sidoarjo pada tahun 2019 ini merupakan daerah yang tergolong rawan. Jika ditinjau dari topografinya, daerah yang tergolong cukup rawan ini merupakan daerah yang memiliki bangunan atau atap rumah yang kurang kokoh. Hal ini dapat diketahui dari data terdampak bahwa terdampak rumah rusak terbesar yaitu 507 rumah.

Selain 5 daerah tersebut (Tulung Agung, Probolinggo, Bojonegoro, Nganjuk dan Sidoarjo) merupakan daerah yang aman. Jika dilihat dari data terdampak 2019 selain 5 daerah tersebut jarang terjadi bencana angin puting beliung. Namun sesekali terjadi angin kencang sehingga terdapat beberapa yang dirugikan. Jadi, walaupun daerah tersebut tergolong aman namun harus tetap waspada akan bencana angin puting beliung.

BAB V

PENUTUP

5.1 Simpulan

Berdasarkan analisis yang telah dilakukan maka diperoleh beberapa kesimpulan sebagai berikut:

1. Berdasarkan metode yang digunakan yaitu K-Means, didapatkan hasil *cluster* dengan daerah yang tergolong rawan dalam hal terdampak pada variabel korban dan kerusakan rumah adalah Kabupaten Tulung Agung karena daerah tersebut merupakan jalur angin sehingga sering dilewati oleh angin monsun barat yang menyebabkan terjadinya angin kencang atau puting beliung. *Cluster* dengan kategori sangat rawan terdapat satu anggota yaitu Sidoarjo. Daerah yang tergolong cukup rawan terdapat 3 daerah yaitu Probolinggo, Nganjuk dan Bojonegoro. Daerah yang tergolong aman terdapat 33 anggota dimana seluruh kabupaten dan kota di Provinsi Jawa Timur kecuali Sidoarjo, Tulung Agung, Probolinggo, Nganjuk, dan Bojonegoro.
2. Hasil uji evaluasi *cluster* K-Means dengan *silhouette*, *davies bouldin*, dan *purity* didapatkan jumlah *cluster* yang optimal sebanyak 4 *cluster*. Dengan *silhouette coefficient* sebesar 0.9116, *davies bouldin index* sebesar 0.3633, dan *purity* sebesar 1. *Silhouette coefficient* memiliki nilai yang cukup tinggi untuk hasil yang optimal, hal ini ditinjau dari minimum jarak dalam *cluster* dan antar *cluster*. *Purity* memiliki nilai yang mirip dengan *silhouette coefficient*, jika mendekati 1 maka *cluster*

- Prilianti, Kestrelia Rega, and Kata Kunci. 2014. "Aplikasi Text Mining Untuk Automasi Penentuan Tren Topik Skripsi Dengan Metode K-Means Clustering." *Cybermatika* 2(1): 1–6. <http://www.mendeley.com/research/aplikasi-text-mining-untuk-automasi-penentuan-tren-topik-skripsi-dengan-metode-kmeans-clustering>.
- Renard, Florent. 2017. "Flood Risk Management Centred on Clusters of Territorial Vulnerability." *Geomatics, Natural Hazards And Risk* 8(2): 525–43. <https://doi.org/10.1080/19475705.2016.1250111>.
- Rohmawati, Nurul, Sofi Defiyanti, and Mohamad Jajuli. 2015. "Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa." *Jurnal Ilmiah Teknologi Informasi Terapan* 1(2): 62–68.
- Saikhu, Ahmad, and Yoga Bhagawad Gita. 2013. "Implementasi Deteksi Outlier Pada Algoritma Hierarchical Clustering." : 45–50.
- Sari, Nilam Novita. 2010. *Pola Spasial Kerentanan Bencana Alam (Studi Kasus Kabupaten Cianjur)*.
- Sari, Nilam Novita. 2018. 15 Universitas Islam Indonesia *K-Affinity Propagation (K-Ap) Clustering Untuk Klasifikasi Gempa Bumi*. Yogyakarta.
- Sirat, Edo Fadila, Budi Darma Setiawan, and Fatwa Ramdani. 2018. "Analisis Perbandingan Algoritme K-Means Dan Isodata Untuk Klasterisasi Data Kejadian Titik Api Di Wilayah Sumatera Pada Tahun 2001 Hingga 2014." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya* 2(11): 5105–12.
- Sulastri, Heni, and Acep Irham Gufroni. 2017. "Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia." 02: 299–305.
- Susanto, B. (2014). *Text and Web Mining*. FTI UKDW.
- Swindiarto, Victory Tyas P. 2018. "Integration of Fuzzy C-Means Clustering and Topsis (Fcm-Topsis) with Silhouette Analysis for Multi Criteria Parameter Data." *2018 International Seminar on Application for Technology of Information and Communication*: 463–68.
- Tanesib, Jehunias, Universitas Nusa Cendana, Nuryanti Yanti, and Universitas Nusa Cendana. 2018. "Pemetaan Daerah Rawan Banjir Dengan Penginderaan Jauh Dan Sistem Informasi Geografis Di Kecamatan Kupang Timur Kabupaten Kupang Provinsi Nusa Tenggara Timur." (August).
- Tanzil Furqon, Muhammad, and Lailil Muflikhah. 2016. "Clustering the Potential Risk of Tsunami Using Density-Based Spatial Clustering of Application With Noise (Dbscan)." *Journal of Enviromental Engineering and Sustainable Technology* 3(1): 1–8.

- Triana, Dessy, Tb Sofwan Hadi, and Muhammad Kamil Husain. 2017. "Mitigasi Bencana Melalui Pendekatan Kultural Dan Struktural." *Seminar Nasional XII "Rekayasa Teknologi Industri dan Informasi 2017"*.
- Yana, Mira Suci, Lathifah Setiawan, Elvitra Mutia Ulfa, and Asep 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.
- Yulian, Eko. 2018. "Text Mining Dengan K-Means Clustering Pada Tema LGBT Dalam Arsip Tweet Masyarakat Kota Bandung." *Jurnal Matematika "MANTIK"* 4(1): 53–58.
- Yuniarti, Mazna. 2018. "Analisis Kejadian Puting Beliung Di Indonesia Menggunakan Spatial Autoregressive (SAR), Clustering Average Linkage, Dan Pemetaan Berbasis WEBGIS." *Director* 15(2): 2017–19.