

Clustering Tingkat Kemiripan Curah Hujan di Indonesia Berdasarkan Provinsi Menggunakan Metode Hierarchical Clustering dan GeoMap

Yahya Supit Runtulalo^{1*}, Daniel H. F. Manongga²

Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia

*e-mail *Corresponding Author*: 672019314@student.uksw.edu

Abstract

Indonesia's geographical location has a big influence on the usual rainfall patterns. BMKG data summarized in the 2023 Indonesian statistical data by the Central Statistics Agency, notes that Indonesia will often experience rain from around August to February. Rainfall is an important factor in determining the climate of an area and has a significant impact on various sectors of life, such as agriculture, water resource management and disaster mitigation. The aim of the research is to find out which provinces in Indonesia have similar levels of low, medium and high rainfall characteristics using the Hierarchical Clustering method and GeoMap visualization. After processing the data, 3 cluster levels were obtained, namely cluster 1 with 8 members in areas with low rainfall levels, cluster 2 with 3 members in areas with high rainfall, and cluster 3 with 22 members in areas with moderate rainfall. Then the test was evaluated using the Silhouette Coefficient which produced cluster 1 (medium), cluster 2 (strong), and cluster 3 (medium).

Keywords: *Rainfall; Hierarchical Clustering; GeoMap*

Abstrak

Letak Geografis Indonesia mempunyai pengaruh besar terhadap pola curah hujan yang biasa terjadi. Data BMKG yang dirangkum dalam data statistik Indonesia 2023 oleh Badan Pusat Statistik, dicatat bahwa Indonesia akan kerap diguyur hujan sekitar bulan Agustus sampai Februari. Curah hujan merupakan faktor penting dalam menentukan iklim suatu daerah dan memiliki dampak yang signifikan terhadap berbagai sektor kehidupan, seperti pertanian, pengelolaan sumber daya air, dan mitigasi bencana. Adapun yang menjadi tujuan dalam penelitian untuk mengetahui provinsi di Indonesia yang memiliki tingkat kemiripan karakteristik curah hujan rendah, sedang, dan tinggi menggunakan metode *Hierarchical Clustering* dan visualisasi GeoMap. Setelah melakukan olah data, didapatkan 3 tingkat cluster yaitu cluster 1 dengan 8 anggota wilayah dengan tingkat curah hujan rendah, *cluster 2* dengan 3 anggota wilayah dengan curah hujan tinggi, dan *cluster 3* dengan 22 anggota wilayah dengan curah hujan sedang. Kemudian di evaluasi pengujian menggunakan *Silhouette Coefficient* yang menghasilkan *cluster 1 (medium)*, *cluster 2 (strong)*, dan *cluster 3 (medium)*.

Kata kunci: *Curah Hujan; Hierarchical Clustering; GeoMap*

1. Pendahuluan

Indonesia merupakan salah satu negara tropis di dunia dengan dua musim setiap tahunnya, yaitu musim kemarau dan musim hujan. Letak Geografis Indonesia sangat berpengaruh terhadap pola curah hujan yang berlangsung sering setiap tahunnya. Data BMKG yang dirangkum dalam data Statistik Indonesia 2023 oleh Badan Pusat Statistik, dicatat bahwa Indonesia akan kerap diguyur hujan sekitar bulan Agustus sampai Februari. Adapun curah hujan di dalam negeri tercatat rata-rata sebesar 2.898 milimeter per tahun hingga tahun 2022 [1]. Curah hujan merupakan faktor penting dalam menentukan iklim suatu daerah dan memiliki dampak yang signifikan terhadap berbagai sektor kehidupan, seperti pertanian, pengelolaan sumber daya air, dan mitigasi bencana. Curah hujan menjadi salah satu parameter penting yang perlu dipahami dan dianalisis dengan baik [2].

Seiring dengan perkembangan teknologi dan ketersediaan data cuaca yang semakin baik, penelitian tentang curah hujan di Indonesia terus mengalami peningkatan tiap tahunnya.

Namun, belum banyak penelitian yang melakukan pengelompokan tingkat kemiripan karakteristik curah hujan di Indonesia berdasarkan provinsi yang ada secara komprehensif. Adapun penelitian sebelumnya oleh Andi [3] yang membahas tentang pemodelan curah hujan di Kota Pontianak menggunakan metode Kuadrat Terkecil non-Linear (*Least Square non Linear*). Pada penelitian yang dilakukan penggunaan metode yang dipilih kurang tepat karena penggunaan data curah hujan pada metode tersebut menggunakan data spasial yang memiliki unsur waktu dalam datanya sehingga, dibutuhkan perlakuan khusus maka data tersebut tidak dapat diproses langsung dengan metode Kuadrat Terkecil non-Linear (*Least Square non-Linear*). Adapun penelitian lainnya yang dilakukan oleh Made dkk [4] yang menganalisis tentang curah hujan dengan metode *Singular Vector Decomposition* dari penelitian tersebut hanya berfokus pada pola curah hujan sehingga analisis yang dihasilkan tidak informatif dan jelas.

Berdasarkan penelitian terdahulu yang dilakukan, diperlukannya sebuah analisis baru untuk membantu melakukan pengelompokan atau *clustering* terhadap data curah hujan menggunakan metode *Hierarchical Clustering*. Metode *Hierarchical Clustering* merupakan teknik klasifikasi yang memungkinkan pengelompokan data berdasarkan tingkat kemiripan atau atribut yang dimiliki serta perbedaan antar objek. Terdapat dua metode dalam melakukan proses *Hierarchical Clustering* yaitu, *Agglomerative* (Penggabungan) pada metode ini, setiap objek dianggap sebagai sebuah kelompok terpisah. Algoritma akan menggabungkan dua kelompok yang paling dekat berdasarkan jarak atau kesamaan antara kelompok, dan proses ini akan berlangsung hingga semua objek digabungkan menjadi satu kelompok utama yang hasil akhirnya menjadi struktur hirarkis dari cluster dan *Divisive* (Pemisahan): metode ini berbanding terbalik dengan metode *Agglomerative*, yang dimana dimulai dengan satu kelompok besar yang berisi semua objek data. Algoritma ini secara berulang memisahkan kelompok menjadi kelompok yang lebih kecil berdasarkan tingkat perbedaan atau jarak antara objek-objeknya. Metode *Divisive* juga menghasilkan struktur hirarkis, tetapi dalam bentuk yang berbeda dengan metode *Agglomerative* [5].

Hasil akhir dari algoritma *Hierarchical Clustering* adalah dendrogram, berupa representasi visual dari struktur hirarkis yang menunjukkan bagaimana kelompok-kelompok lebih kecil dan lebih besar saling berhubungan. Dendrogram juga dapat digunakan untuk memahami hubungan antara kelompok-kelompok serta untuk memutuskan pada jumlah dan komposisi kelompok yang lebih optimal dalam analisis lebih lanjut [6]. Adapun yang menjadi tujuan dalam penelitian ini adalah untuk mengetahui provinsi mana saja di Indonesia yang mempunyai tingkat persamaan karakteristik curah hujan rendah, sedang dan tinggi dengan menggunakan metode *Hierarchical Clustering* dan visualisasi *GeoMap*, serta mengetahui hasil *clustering* mengenai tingkat curah hujan di wilayah provinsi di Indonesia menggunakan metode *Hierarchical Clustering*.

Sehingga melalui penelitian ini dapat memberikan manfaat berupa informasi hasil *clustering* provinsi di Indonesia yang memiliki curah hujan rendah, sedang, dan tinggi yang dapat menjadi upaya preventif terjadinya bencana alam yang diakibatkan oleh curah hujan dan menjadi sebuah pertimbangan dalam berbagai kegiatan serta aktivitas dalam masyarakat yang memiliki keterkaitan dengan curah hujan serta, hasil informasi nantinya diharapkan akan sangat berguna bagi pemerintah, lembaga riset, dan pihak terkait lainnya dalam pengambilan keputusan terkait perencanaan pembangunan, pengurangan risiko bencana, pengelolaan sumber daya air yang disebabkan oleh curah hujan ekstrim, serta mitigasi risiko bencana terkait curah hujan.

2. Tinjauan Pustaka

Pada Penelitian terdahulu yang berjudul "Pengaruh Siklon Tropis Cempaka Terhadap Curah Hujan Harian Di Wilayah Jawa dan Madura" menjelaskan tentang analisis yang dilakukan untuk mengidentifikasi pola curah hujan yang ada di Jawa dan Madura dengan menggunakan metode *Clustering Hierarchical*. Hasilnya menunjukkan bahwa ada empat karakteristik kelompok curah hujan yang paling erat terkait dengan siklon tropis Cempaka. Analisis pola curah hujan dari keempat kelompok tersebut menunjukkan bahwa siklon tropis Cempaka memiliki dampak yang signifikan terhadap jumlah curah hujan yang terjadi di Jawa dan Madura [7] dari penelitian tersebut hanya berfokus pada satu daerah saja sehingga mempunyai perbandingan curah hujan yang kecil.

Penelitian terdahulu yang berjudul "Pewilayahan Curah Hujan di Kabupaten Indramayu dengan Metode Gerombol (berdasarkan Data Median Tahun 1980-2000) menggunakan metode *divisive hierarchical* dan metode *agglomerative hierarchical* untuk mengelompokkan data median

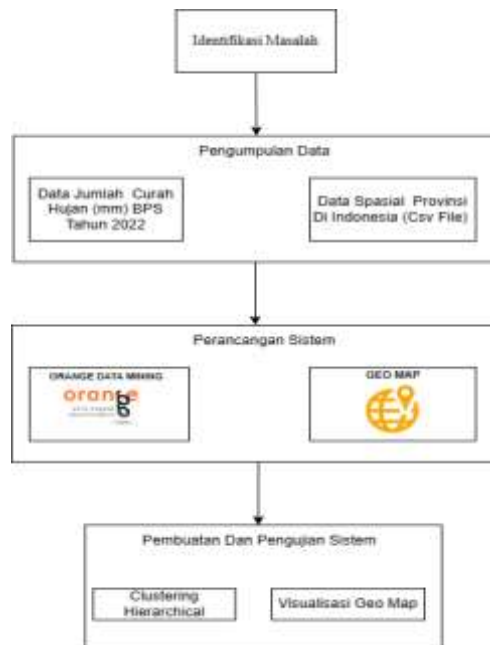
curah hujan di 27 stasiun penakar hujan di Indramayu. Selain itu, jarak dihitung menggunakan jarak Euclid dan hubungan lengkap. Hasil penelitian menunjukkan bahwa ada tiga wilayah baru. Wilayah 3 mengalami curah hujan tertinggi pada bulan Januari berdasarkan pola dan sebaran di wilayah tersebut. Wilayah 2, yang terletak di sebelah barat daya dan selatan dan berbatasan dengan Kabupaten Sumedang dan Majalengka, mengalami curah hujan tahunan tertinggi [8].

Penelitian terdahulu yang berjudul “ Pemodelan Statistical Downscaling Dengan Peubah Dummy Berdasarkan Teknik *Cluster* Hierarki Dan Nonhierarki Untuk Pendugaan Curah Hujan “menggunakan Metode hierarkis dan non-hierarkis dalam penelitian ini untuk memprediksi skala curah hujan lokal berdasarkan sirkulasi global. Penelitian ini menggunakan model cluster (hubungan rata-rata, hubungan median, dan hubungan lingkungan) untuk membuat kelompok data curah hujan. Tujuan dari penelitian ini adalah untuk memperkirakan jumlah curah hujan lokal di Kabupaten Pangkep, yang merupakan daerah penghasil garam di Sulawesi Selatan. Studi ini menemukan bahwa metode *cluster* non-hierarki *K-means* memiliki hasil pemodelan dan validasi model RKU terbaik [9]. Penelitian ini berfokus hanya pada prakiraan pola curah hujan yang ada di 1 daerah provinsi saja sehingga tidak mendapat suatu perbandingan yang lebih luas dan signifikan .

Berdasarkan penjelasan sebelumnya, maka dilakukan penelitian yang baru untuk membantu pengelompokan data tingkat kemiripan curah hujan yang ada di Indonesia pada tahun 2022 berdasarkan provinsi menggunakan algoritma *Hierarchical clustering*, Hal ini untuk mengetahui klasifikasi setiap data tingkat kemiripan curah hujan mulai dari rendah, sedang dan tinggi serta hasil yang divisualisasikan melalui gambar. Penelitian ini dilakukan dengan menggunakan *tools* aplikasi Orange dalam menghasilkan *cluster* terhadap karakteristik curah hujan dan hasil cluster akan di evaluasi menggunakan tahap pengujian menggunakan *Silhouette Score* yang ada di dalam *tools Orange*.

3. Metodologi

Penelitian ini mempunyai beberapa tahapan dan alur metode yang akan dilakukan untuk menyelesaikan masalah. Permasalahan penelitian ini mulai dari bagaimana melakukan suatu klasterisasi atau pengelompokan pada data curah hujan dan menampilkan hasil visualisasinya dengan menggunakan *GeoMap*. Adapun tahapan alur selama penelitian dapat dilihat pada Gambar 1.



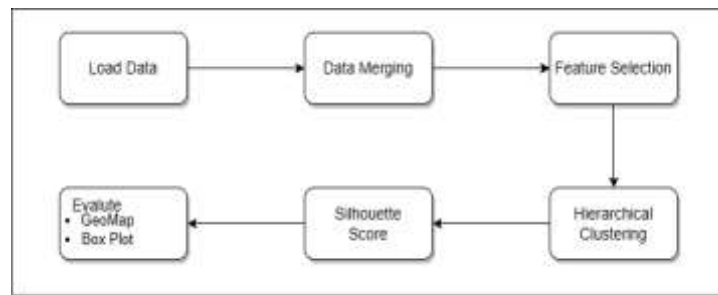
Gambar 1 Metode Penelitian

Pada Identifikasi permasalahan akan dilakukan pemahaman terkait permasalahan dengan tujuan memahami tingkat curah hujan yang ada di Indonesia berdasarkan provinsi yang ada. Batasan penelitian yang dilakukan mengamati tingkat curah hujan dari tinggi, sedang, dan

rendah di provinsi yang terdapat di Indonesia. proses pengumpulan data sesuai dengan penelitian yang akan dilakukan. Adapun data yang akan digunakan diperoleh melalui *website* Badan Pusat Statistik yaitu data jumlah curah hujan (*mm*) yang terjadi di provinsi per tahun 2022. Data berikutnya juga data spasial provinsi yang ada di Indonesia yang diperoleh melalui *website* Indonesia Geospasial. Proses *clustering* dilakukan menggunakan aplikasi *Orange* dengan bersifat *open source* dan *tools* yang ditemukan dalam aplikasi ini sehingga akan mendukung proses untuk analisis data. Setelah diperoleh hasil dari proses *clustering*, maka akan dilakukan proses visualisasi dengan penggunaan *GeoMap* dimana *tools* tersebut ditemukan dalam aplikasi *Orange*. Tahap terakhir yaitu tahap pembuatan dan pengujian pada sistem yang dilakukan *clustering* atau pengelompokkan. hasil dari *clustering* akan dilakukan visualisasi ke *GeoMap*. metode yang digunakan dari hasil *clustering* yaitu *Agglomerative Hierarchical Clustering* dimana Tahap pelaksanaan atau implementasi yang dilakukan saat memproses atau mengelompokkan dataset yang telah terkumpul dan diproses dengan pendekatan algoritma *Agglomerative Hierarchical Clustering* adalah dengan *Complete Linkage*. untuk formula *Complete Linkage* sebagai berikut

$$d_{(UV)W} = \max(d_{UW}, d_{VW}) \dots \dots \dots (1)$$

proses visualisasi digunakan data spasial yang mempunyai atribut titik koordinat pada letak geografis setiap provinsi di Indonesia. visualisasi akan dimulai dengan melakukan pengambilan hasil dari klasterisasi pada proses *clustering* yang telah dilakukan sebelumnya. pada *GeoMap* akan ditampilkan peta dengan hasil berupa warna yang memiliki perbedaan didasarkan pada hasil *clustering*.



Gambar 2. Arsitektur Penelitian

Gambar 2 adalah alur sebuah kerangka kerja proses penelitian yang diawali dari memuat data yang telah didapatkan sebelumnya, melakukan proses penggabungan dataset atau *merging data* dari kedua data untuk nantinya dilakukan proses klasterisasi. Setelah itu, dilakukan *feature selection* untuk menentukan atau memilih atribut kolom dari dataset yang akan digunakan untuk memudahkan proses *clustering* yang akan dilakukan serta dapat memberikan hasil yang efisien. Selanjutnya, dilakukan *clustering* dengan algoritma *Agglomerative Hierarchical Clustering* menggunakan pendekatan metode *complete linkage* untuk memperoleh hasil *clustering* dari penelitian yang dilakukan.

setelah melakukan *clustering*, dilakukan pengujian atau pengecekan hasil *clustering* menggunakan *Silhouette Coefficient* untuk melihat hasil atau tingkat keberhasilan proses *clustering* data penelitian menggunakan *Hierarchical clustering*. Nilai *Silhouette Coefficient* berkisar antara -1 dan 1. Semakin mendekati 1, maka semakin bagus dan tepat pengelompokkan yang didapatkan dari hasil *clustering*. Berikut adalah nilai *Silhouette Coefficient* berdasarkan Kauman dan Rousseeuw [10] dapat dilihat pada tabel 1.

Tabel 1. Hasil Structure Nilai *Silhouette Coefficient* [10]

Nilai <i>Silhouette Coefficient</i>	Hasil Structure
$0.7 \leq 1$	Strong Structure
$0.5 \leq 0.7$	Medium Structure
$0.25 \leq 0.5$	Weak Structure
≤ 0.25	No Structure

Proses terakhir setelah melakukan *clustering* yaitu evaluasi hasil dengan visualisasi dari hasil *clustering* menggunakan *GeoMap* yang ditampilkan dalam bentuk gambar peta geografis berdasarkan data spasial agar mempermudah pemahaman pada hasil dari penelitian.

4. Hasil dan Pembahasan

Pada penelitian ini dilakukan pengumpulan data curah hujan melalui website resmi Badan Pusat Statistik pada tahun 2022 [11]. Data tersebut dimasukkan ke dalam format excel dan ditampilkan seperti yang ada pada tabel 2.

Tabel 2. Data Curah Hujan(*mm*) Tahun 2022 [11]

Provinsi	Volume_hujan
Aceh	1834,1
Bali	2834,4
Banten	2877,2
Bengkulu	4518,5
DI Yogyakarta	3161,2
DKI Jakarta	2136,3
Gorontalo	2082,8
Jambi	3265,6
Jawa Barat	3505,4
Jawa Tengah	2643,9
Jawa Timur	3059,7
Kalimantan Barat	3040,6
Kalimantan Selatan	2738,1
Kalimantan Tengah	3259,1
Kalimantan Timur	3313,7
Kepulauan Bangka Belitung	3026,7
Kepulauan Riau	2590,6
Lampung	2099,1
Maluku	3766
Maluku Utara	2913,9
Nusa Tenggara Barat	2649,9
Nusa Tenggara Timur	2183
Papua	2576,3
Papua Barat	1549
Riau	3231,1
Sulawesi Barat	2395,8
Sulawesi Selatan	4650,9
Sulawesi Tengah	879,4
Sulawesi Tenggara	2695,4
Sulawesi Utara	3529,8
Sumatera Barat	4950,5
Sumatera Selatan	3250,2
Sumatera Utara	3269,5

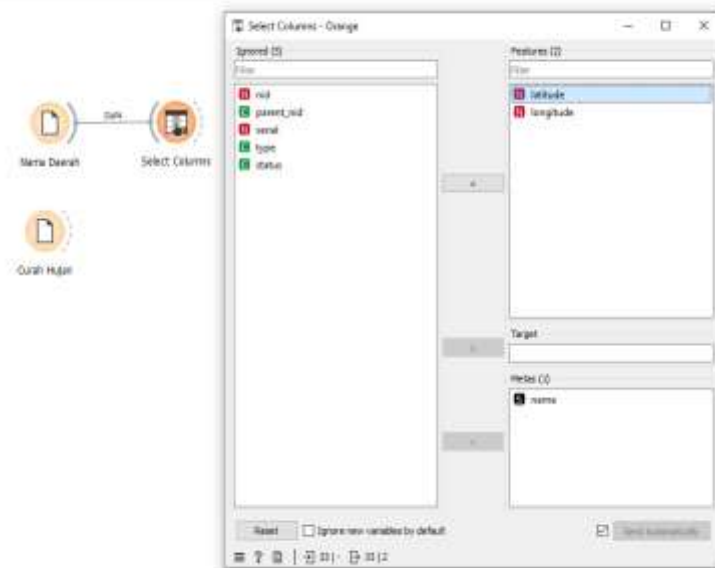
Data berikutnya yang digunakan juga yaitu data spasial yang diperoleh melalui Indonesia -geospasial. Pada tampilan data tersebut akan terdapat berbagai atribut dan kolom yang terdiri dari identitas setiap wilayah provinsi seperti yang ada di Tabel 3.

Tabel 3. Data Wilayah

Id	name	Serial	latitude	Longitude	status
1.	Aceh	11	4695135	967493993	1
2.	Bali	51	-84095178	115188916	1

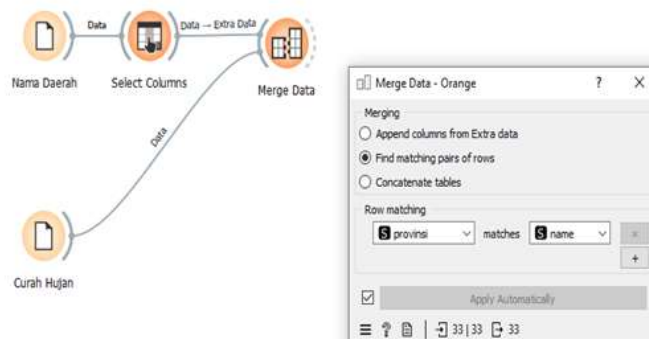
Id	name	Serial	latitude	Longitude	status
3.	Banten	36	-64058172	1060640179	1
4.	Bengkulu	17	-35778471	1023463875	1
5.	DI Yogyakarta	34	-78753849	1104262088	1
6.	DKI Jakarta	31	-6211544	106845172	1
7.	Gorontalo	75	6999372	1224467238	1
8.	Jambi	15	-14851831	1024380581	1
9.	Jawa Barat	32	-7090911	107668887	1
10	Jawa Tengah	33	-7150975	1101402594	1
11	Jawa Timur	35	-75360639	1122384017	1
12	Kalimantan Barat	61	-2787808	1114752851	1
13	Kalimantan Selatan	63	-30926415	1152837585	1
14	Kalimantan Tengah	62	-16814878	1133823545	1
15	Kalimantan Timur	64	16406296	116419389	1
	Kepulauan Bangka				
16	Belitung	19	-27410513	1064405872	1
17	Kepulauan Riau	21	39456514	1081428669	1
18	Lampung	18	-45585849	1054068079	1
19	Maluku	81	-32384616	1301452734	1
20	Maluku Utara	82	15709993	1278087693	1
	Nusa Tenggara				
21	Barat	52	-86529334	1173616476	1
	Nusa Tenggara				
22	Timur	53	-86573819	1210793705	1
23	Papua	94	-4269928	1380803529	1
24	Papua Barat	91	-13361154	1331747162	1
25	Riau	14	2933469	1017068294	1
26	Sulawesi Barat	76	-28441371	1192320784	1
27	Sulawesi Selatan	73	-36687994	1199740534	1
28	Sulawesi Tengah	72	-14300254	1214456179	1
29	Sulawesi Tenggara	74	-414491	122174605	1
30	Sulawesi Utara	71	6246932	1239750018	1
31	Sumatera Barat	13	-7399397	1008000051	1
32	Sumatera Selatan	16	-33194374	103914399	1
33	Sumatera Utara	12	21153547	995450974	1

Data tersebut akan melalui *select column* yang bertujuan untuk memilih atau menentukan kolom atribut yang diperlukan untuk identitas suatu wilayah. Pada proses ini akan dilakukan penghapusan kolom yang tidak digunakan, sehingga memori yang digunakan dalam pemrosesan data lebih efisien dan optimal [12]. Proses *clustering* menggunakan algoritma *Hierarchical Clustering* akan dilakukan pada aplikasi yang bernama *Orange Data Mining*.



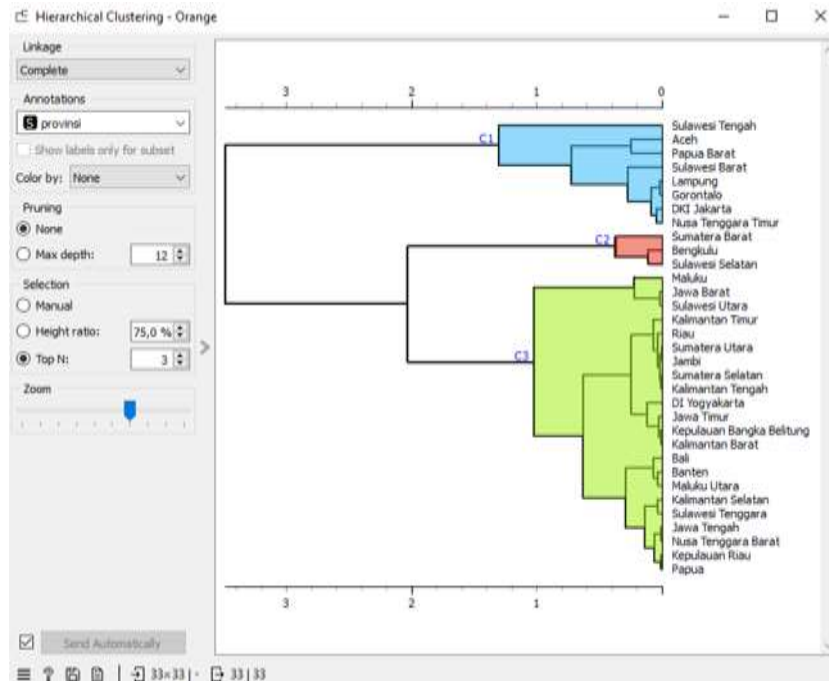
Gambar 3. Select Columns Atribut

Pada gambar 3 diatas, langkah pertama yang dilakukan dalam pemrosesan data dalam aplikasi Orange Data Mining yaitu *load* dataset yang sudah didapatkan sebelumnya, kemudian akan dilakukan proses *select columns* pada data wilayah yang dimana atribut data yang dipilih untuk *clustering* hanya data atribut mengenai nama daerahnya beserta *Latitude* dan *Longitudenya*, data spasialnya disini hanya digunakan untuk menunjukkan lokasi dena dari sebuah wilayah di Indonesia.



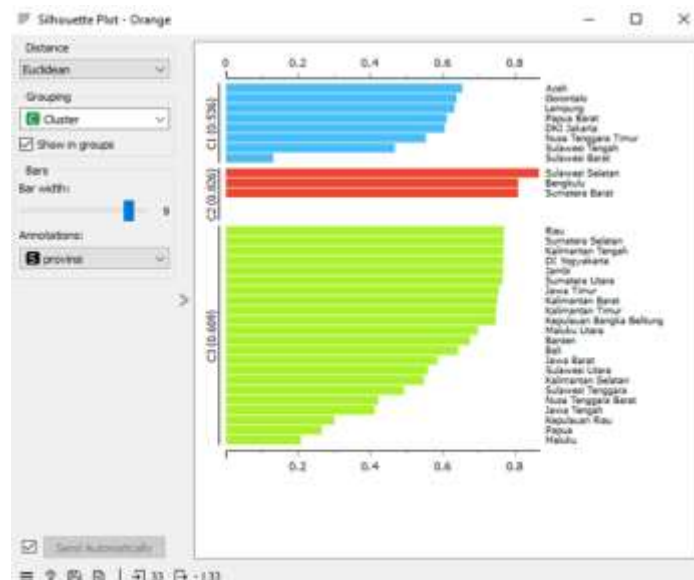
Gambar 4. Merge Data

Penggabungan data dari kedua kesamaan data dalam hal ini menggabungkan atribut yang sama dimiliki oleh kedua data tersebut yaitu atribut data Provinsi dan Name. standarisasi nilai data dilakukan menggunakan *Distances* dari data yang diperoleh sebelumnya sehingga jarak antar nilai yang diperoleh tidak terlalu bermacam-macam pada proses *clustering*. Proses awal *clustering* dilakukan dengan menggunakan model analisis perhitungan *complete linkage* (jarak terjauh) dan Annotations (anotasi) clusternya berdasarkan provinsi. hasil *clustering* akan ditampilkan pada Gambar 5. Keterangan dari hasil ketiga cluster yaitu pada cluster 1 tingkat curah hujan rendah, cluster 2 tingkat curah hujan sedang, dan cluster 3 tingkat curah hujan tinggi. Berdasarkan hasil dari *clustering* dengan penggunaan metode algoritma *Hierarchical clustering* yang ditampilkan dalam bentuk *Hierarki*, cluster 1 mempunyai 8 anggota, cluster 2 mempunyai 3 anggota, dan cluster 3 mempunyai 22 anggota [13].



Gambar 5. Hasil Proses Clustering

Proses *Silhouette Coefficient* untuk melakukan evaluasi dari hasil pengujian terhadap algoritma Hierarchical clustering dalam bentuk representasi grafis, serta melihat performa jarak data dalam satu *cluster* yang diperoleh dalam hasil *clustering* dengan metode *Hierarchical Clustering*. Dengan menggunakan model *Distance euclidean* sebagai analisis jarak data seperti yang ada pada Gambar 6, dihasilkan nilai tertinggi ada di cluster 2 dengan nilai *Silhouette score* sebesar 0,826. cluster 1 mendapatkan hasil 0,536 (*Medium*) dengan 8 anggota, cluster 2 mendapatkan hasil 0,826 (*Strong*) dengan 3 anggota dan cluster 3 dengan hasil 0,609 (*Medium Structure*) dengan 22 anggota [13], [14].



Gambar 6. Silhouette Score

Setelah diperoleh hasil dari clustering, dilakukan proses *merge* data seperti yang di tampilkan pada gambar 7 dari data hasil cluster yang didapatkan sebelumnya dengan dataset

wilayah yang di masukkan sebelumnya di awal untuk mempermudah proses visualisasi menggunakan *GeoMap*.

	name	Cluster	latitude	longitude	volume_rujan
1	Aceh	C1	-4.89513	98.7494	1234.1
2	Sumatera Utara	C3	2.11335	99.5451	3266.5
3	Sumatera Barat	C2	-0.73994	100.8	4930.5
4	Jawa	C3	0.29347	101.707	3331.1
5	Jambi	C3	-1.48598	103.438	3265.6
6	Sumatera Selatan	C3	-3.21944	103.914	3230.2
7	Bengkulu	C2	-3.77785	102.348	4518.5
8	Lampung	C1	-4.55836	105.407	2099.1
9	Kepulauan Riau	C1	-2.74129	106.441	3026.7
10	Kepulauan Riau	C2	3.94965	108.143	2590.6
11	DKI Jakarta	C1	-6.21194	106.845	2136.3
12	Jawa Barat	C1	-7.30081	107.669	3924.4
13	Jawa Tengah	C2	-7.53087	110.14	2642.9
14	DI Yogyakarta	C2	-7.87328	110.426	3181.2
15	Jawa Timur	C3	-7.53606	112.238	3696.7
16	Banten	C3	-6.42582	106.064	2877.2
17	Bali	C3	-8.49552	115.189	2834.4
18	Nusa Tenggara...	C3	-8.65283	117.362	2640.0
19	Nusa Tenggara...	C1	-8.85738	121.079	2183.0
20	Kalimantan Barat	C3	-0.278781	111.475	3940.6
21	Kalimantan Ten...	C2	-1.68149	113.382	3236.1
22	Kalimantan Sel...	C3	-0.99284	115.284	2736.1
23	Kalimantan Ten...	C3	1.64083	119.418	3313.7
24	Sulawesi Utara	C3	0.626889	123.975	1529.8
25	Sulawesi Tengah	C1	-1.43029	121.446	876.4
26	Sulawesi Selatan	C2	-2.9688	119.874	4830.9
27	Sulawesi Tenggara	C3	-4.54691	122.175	2895.4
28	Gorontalo	C1	0.699937	122.447	2082.8
29	Sulawesi Barat	C1	-2.84814	119.732	2795.8
30	Maluku	C3	-1.23846	130.145	3766.0
31	Maluku Utara	C3	1.371	127.809	2813.9
32	Papua Barat	C1	-1.33612	135.175	1349.0
33	Papua	C3	-4.26989	138.08	2576.3

Gambar 7. Data Hasil Clustering

Visualisasi menggunakan *GeoMap* dilakukan agar dapat mempermudah pemahaman terkait hasil clustering dan memberikan hasil visualisasi yang lebih menarik. Proses visualisasi menggunakan *widget* yang ada dalam *tools Orange* dan berdasarkan dataset pada hasil clustering dan menggunakan atribut data *latitude* dan *longitude* sebagai titik koordinat suatu wilayah. bentuk peta seperti yang ditunjukkan pada Gambar 8 dan Gambar 9.



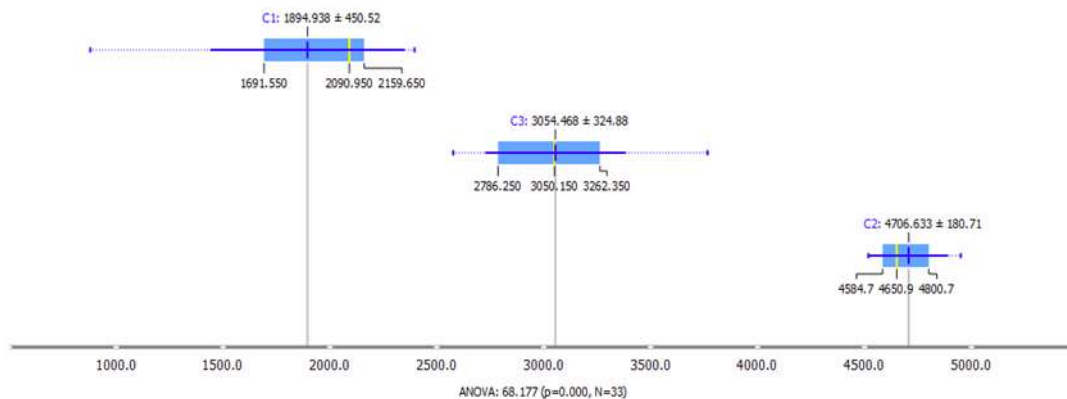
Gambar 8. Visualisasi *GeoMap*

Visualisasi kedalam peta akan menampilkan identitas mengenai tingkat intensitas curah hujan yang ada di beberapa lokasi titik koordinat yang sudah terbagi menjadi 3 *cluster* dan dibedakan dalam bentuk pewarnaan titik setiap *cluster* dalam peta.



Gambar 9. Keterangan Visualisasi GeoMap

Saat di-klik pada bagian objek hasil *clustering*, maka akan muncul informasi tentang objek tersebut, seperti yang ditunjukkan pada Gambar 9. kemudian melakukan visualisasi data menggunakan *boxplot* untuk mengetahui secara terperinci karakteristik dari *cluster* yang dihasilkan sebelumnya dan untuk mengetahui juga bagaimana penyebaran suatu data yang sudah di *clustering*.



Gambar 10. Box Plot

Hasil dari menggunakan *Boxplot* ditunjukkan pada gambar 10, memberikan hasil karakteristik setiap *cluster* berbeda-beda, dimana terdapat cluster 1 berada di posisi tingkat curah hujan rendah yang memiliki range data yaitu dari 1691,5 sampai 2159,6. kemudian *cluster* 2 berada di posisi tingkat curah hujan paling tinggi yang memiliki range data curah hujan dalam *clusternya* yaitu dari 4584,7 sampai 4800,7 kemudian *cluster* 3 berada di posisi tingkat curah hujan sedang yang memiliki range data rata-rata curah hujan dalam *clusternya* berkisar dari 2786,2 sampai 3262,3 [15].

5. Simpulan

Hasil dari penelitian ini menggunakan metode *Hierarchical Clustering* dan visualisasi *GeoMap* dapat menunjukkan sebuah informasi awal mengenai tingkat kesamaan karakteristik curah hujan yang ada di provinsi Indonesia yang memiliki curah hujan rendah, sedang, dan tinggi sebagai acuan awal pengambilan keputusan bagi pemerintah, lembaga riset dan pihak terkait lainnya dalam perencanaan pembangunan, pengurangan risiko bencana, pengelolaan sumber daya air yang disebabkan oleh curah hujan ekstrim, serta mitigasi risiko bencana terkait curah hujan. Dari analisis hasil clustering menggunakan metode *Hierarchical Clustering* yang telah dilakukan evaluasi hasil menggunakan metode pengujian *Silhouette Coefficient*, didapatkan hasil yang sangat baik (*Strong Structure*) dengan 3 kelompok karakteristik kategori tingkat curah hujan. Adapun kategori yang masuk ke dalam klaster 1 adalah wilayah provinsi dengan tingkat kemiripan curah hujan yang rendah, sehingga klaster 1 adalah wilayah yang jarang terjadi hujan, dan klaster 2 adalah wilayah yang paling sering terjadi hujan, sehingga klaster 2 memiliki tingkat kemiripan curah hujan yang paling tinggi diantara klaster yang lain serta, klaster 3 adalah wilayah yang dengan tingkat kemiripan curah hujan yang sedang, karena merupakan wilayah yang cukup sering terjadi hujan. Klaster 1 mempunyai 8 anggota wilayah, sedangkan klaster 2 mempunyai 3 anggota, dan klaster 3 mempunyai 22 anggota wilayah. Penelitian ini hanya berfokus kepada pengelompokan beserta hasil dari *clustering* karakteristik mengenai tingkat kemiripan curah hujan yang ada di provinsi Indonesia menggunakan tools dari aplikasi Orange, sehingga dalam Pengujian mengenai tingkat akurasi kinerja metode yang digunakan belum diuji dalam penelitian ini, maka dari itu saran untuk penelitian lebih lanjut yaitu diharapkan dapat melakukan perancangan ulang beserta pengujian tingkat akurasi terhadap kinerja metode *Hierarchical Clustering* untuk bisa mendapatkan perhitungan yang informatif terhadap hasil tingkat akurasi kinerja metode dalam proses clustering yang lebih baik dari sebelumnya,

Daftar Referensi

- [1] Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), "Konsep Curah Hujan ",2023. [Online]. Tersedia: <https://www.bmkg.go.id/>. [Diakses: 16 Mei 2023].
- [2] Pratikto, R. O., Damastuti, N., "Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 6, no. 1, pp.13-20, 2021.
- [3] Ihwan Andi, "Pemodelan Curah Hujan Bulanan Berdasarkan Metode Least Square non Linear," *Pontianak: Prodi Fisika FMIPA Universitas Tanjungpura*, 2010.
- [4] Dwi, Made.,Wahyu, Tri., Syahputra, M Rido, "Analisis Curah Hujan di Indonesia untuk Memetakan Daerah Potensi Banjir dan Tanah Longsor dengan Metode Cluster Fuzzy C-Means dan Singular Value Decomposition (SVD)," *Bukit Kotabang: MEGASAINS*, 2014.
- [5] ELAN, D., & Hadi, S, "Penerapan Metode Agglomerative Hierarchical Clustering untuk Klasifikasi Dokumen Skripsi ", *Doctoral dissertation Universitas Bina Darma*, 2020.
- [6] Widyawati, W., Saptomo, W. L. Y., Utami, Y. R. W, " Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan ", *Jurnal Ilmiah SINUS*, vol. 18, no. 1, pp.75-87, 2020.
- [7] Habibie, M. N., Noviati, S., & Harsa Hastuadi, "Pengaruh Siklon Tropis Cempaka Terhadap Curah Hujan Harian di Wilayah Jawa dan Madura," *Meteorologi Dan Geofisika*, vol. 19, no. 1, pp.1-11, 2018.
- [8] Saputro, D. R. S., Mattjik, A. A., Boer, R., Wigena, A. H., & Djuraidah Anik, "Pewilayahan Curah Hujan Di Kabupaten Indramayu Dengan Metode Gerombol (Berdasarkan Data Median Tahun 1980–2000)," *Universitas Negeri Yogyakarta*, 2011.
- [9] Sahriman, S., Kalondeng, A., & Koerniawan Vier, "Pemodelan Statistical Downscaling Dengan Peubah Dummy Berdasarkan Teknik Cluster Hierarki Dan Non-Hierarki Untuk Pendugaan Curah Hujan," *Indonesian Journal of Statistics and Its Applications*, vol. 3, no. 3, pp.295-309, 2019.
- [10] Utami, D. S., & Saputro Dewi Retno Sari, "Pengelompokan data yang Memuat Pencilan dengan Kriteria Elbow dan Koefisien Silhouette (Algoritme K-Medoids)," *Universitas Muhammadiyah Surakarta*, 2018.

-
- [11] Badan Pusat Statistik Indonesia (BPS), “ Statistik Indonesia 2023 “, 2023. [Online]. Tersedia: <https://www.bps.go.id/> [Diakses: 16 Mei 2023].
 - [12] Liebchen, G. A., “ Data cleaning techniques for software engineering data sets “, *PhD diss.*, no. October, 2010.
 - [13] Yulianti, D., Hermanto, T., & Defriani, M., “Analisis Clustering Donor Darah dengan Metode Agglomerative Hierarchical Clustering “, *Resolusi: Rekayasa Teknik Informatika dan Informasi*, vol. 3, no. 6, pp.436-441, 2023.
 - [14] Shahapure, K. R., & Nicholas, C., “Cluster quality analysis using silhouette score “, *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)*, pp. 747-748, 2020. IEEE
 - [15] Hofmann, H., Wickham, H., & Kafadar, K., “value plots: Boxplots for large data “, *Journal of Computational and Graphical Statistics*, vol. 26, no. 3, pp. 469-477, 2017.