

KLASIFIKASI CURAH HUJAN DI INDONESIA MENGGUNAKAN METODE HIERARCHICAL CLUSTERING

Siti Salmah Zakiah¹, Shelly Carista Devibarus², Muhamad Malik Mutoffar³, Sri Kuswayati⁴
salmahzakiah11@gmail.com¹, shellycaristadevi22@gmail.com², malik@sttbandung.ac.id³,
srikuswayati@sttbandung.ac.id⁴
Sekolah Tinggi Teknologi Bandung

ABSTRAK

Klasifikasi curah hujan di Indonesia memiliki peran penting dalam pemahaman pola iklim dan manajemen sumber daya air. Penelitian ini bertujuan untuk mengidentifikasi pola klasifikasi curah hujan berdasarkan variabilitas spasial dan temporal. Dataset curah hujan dari beberapa stasiun cuaca di Indonesia dikumpulkan dan diolah menggunakan perangkat lunak Orange. Metode Hierarchical Clustering digunakan untuk mengelompokkan data curah hujan dari berbagai stasiun cuaca di Indonesia menggunakan platform analisis data Orange. Hierarchical Clustering diterapkan untuk mengelompokkan stasiun-stasiun tersebut berdasarkan pola curah hujan bulanan mereka selama beberapa tahun terakhir. Pemilihan fitur yang tepat, seperti suhu udara, kelembaban, dan tekanan udara, menjadi kunci dalam pengelompokan yang akurat. Hasil analisis menunjukkan bahwa Hierarchical Clustering mampu mengidentifikasi kelompok stasiun cuaca dengan pola curah hujan yang serupa. Klasifikasi curah hujan menghasilkan 5 cluster. Diantaranya Cluster 1 yaitu hujan sangat lebat, cluster 2 hujan lebat, cluster 3 adalah hujan sedang, cluster 4 yaitu hujan ringan, dan cluster 5 adalah tidak hujan. Klasifikasi curah hujan pada provinsi yang ada di Indonesia terdapat 2 provinsi yang masuk ke cluster 1, 5 provinsi yang masuk ke cluster 2, 10 provinsi yang masuk ke cluster 3, 7 provinsi yang masuk ke cluster 4 dan 9 provinsi yang masuk ke cluster 5.

Kata Kunci: Curah Hujan, Hierarchical Clustering, Klasifikasi Curah Hujan

PENDAHULUAN

Indonesia merupakan negara kepulauan terbesar di dunia dengan 17.508 pulau yang dihuni oleh lebih dari 360 suku bangsa. Secara geografis, Indonesia berada diposisi silang karena ada diantara dua benua dan dua samudra. Yaitu benua Asia dan Benua Australia. Untuk Samudranya ada Samudra Pasifik dan Samudra Hindia. Posisi silang ini sangat strategis dan memberikan banyak dampak yang bagus bagi kehidupan masyarakat di tanah air, salah satunya kondisi musim dan iklim. Iklim musim terdiri atas angin muson barat dan timur. Angin muson barat membawa musim hujan yang bertiup pada bulan Oktober-April. Sedangkan angin muson timur membawa musim kemarau yang bertiup pada bulan Mei-September.

Tetapi sekarang ini musim seringkali berubah-ubah. Terdapat beberapa faktor yang mempengaruhi kondisi cuaca. Antara lain kelembapan, suhu, angin, curah hujan, sinar matahari, dan tekanan atmosfer. Faktor tersebut dijadikan atribut dalam memprediksi cuaca yang terjadi diwaktu mendatang. Banyak orang yang akan terbantu dengan adanya prakiraan cuaca. Apalagi untuk pekerjaan dalam bidang pertanian dan kelautan. Prediksi cuaca yang akurat dan cepat dapat memudahkab stakholdier dalam mengelola informasi perubahan cuaca(Muhandis et al, 2021).

Hujan merupakan komponen utama yang mengisi ketersediaan air tawar di bumi. Tinggi rendahnya hujan secara timbal balik akan mempengaruhi kondisi vegetasi, tutupan lahan, dan bahan baku lahan. Oleh karena itu hujan merupakan komponen utama dari unsur iklim, lalu disusul dengan temperatur dan kelembaban.

Selain itu hujan sangat penting untuk keberlangsungan hidup manusia. Untuk pertanian, pengelolaan sumber daya air, dan mitigasi bencana. Karena itu dibutuhkan kajian untuk mengklasifikasi curah hujan berdasarkan provinsi di Indonesia menggunakan metode Hierarchical Clustering, Tujuan dari klasifikasi tersebut adalah memberikan gambaran dasar keberadaan hujan serta karakteristik dari hujan tersebut. Pengklasifikasian data juga dimaksudkan untuk mendapatkan informasi yang sebanyak-banyaknya tentang variasi jumlah cluster.

METODOLOGI

Data yang digunakan dalam penelitian ini adalah data spasial curah hujan tiap provinsi yang ada di Indonesia pada tahun 2020 yang didapat dari github dengan url <https://github.com/Kodesiana/YT-Orange-Data-Mining-Series/tree/master/datasets>. Variable yang digunakan adalah jumlah volume curah hujan yang ada di provinsi Indonesia. Kemudian akan diolah dengan melakukan clustering volume curah hujan berdasarkan provinsi di Indonesia dalam tingkat 5 cluster. Berikut adalah data volume hujan di setiap provinsi.

Tabel 1.

Data Provinsi dan jumlah volume hujan di Indonesia
Sumber: <https://github.com/Kodesiana/YT-Orange-Data-Mining-Series/tree/master/datasets>

Provinsi	Volume_hujan
Aceh	1575
Sumatera Utara	975.9
Sumatera Barat	3548
Riau	2048.3
Jambi	1694.9
Sumatera Selatan	1947.2
Bengkulu	2668.9
Lampung	1628.1
Kepulauan Bangka Belitung	1534.7
Kepulauan Riau	2250.9
DKI Jakarta	2169.5
Jawa Barat	2199.3
Jawa Tengah	1620.7
DI Yogyakarta	2045.5
Jawa Timur	2024.7
Banten	1310.1
Bali	1133.8
Nusa Tenggara Barat	1147.9
Nusa Tenggara Timur	1406
Kalimantan Barat	2757.7
Kalimantan Tengah	2748.4
Kalimantan Selatan	2509.6
Kalimantan Timur	2069.4
Sulawesi Utara	1807
Sulawesi Tengah	460.9

Provinsi	Volume_hujan
Sulawesi Selatan	3382
Sulawesi Tenggara	1589.6
Gorontalo	870.6
Sulawesi Barat	1167.9
Maluku	1987.2
Maluku Utara	913.4
Papua Barat	2844.6
Papua	1265.9

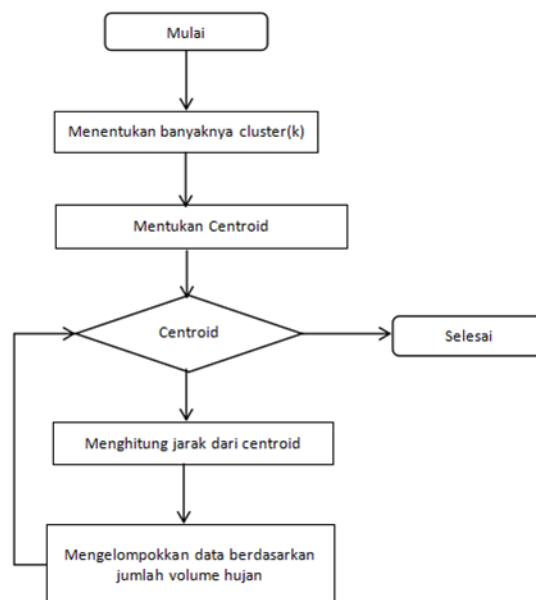
Tabel 2.

Skala Clustering Curah Hujan

Kategori	Curah Hujan
C1	Hujan Sangat Lebat
C2	Hujan Lebat
C3	Hujan Sedang
C4	Hujan Ringan
C5	Tidak Hujan

Analisis Clustering adalah proses membagi sekelompok data menjadi beberapa kelompok, kesamaan satu kelompok data dan lebih besar daripada kelompok data lainnya. Dalam menentukan cluster berdasarkan data yang tersedia, dibutuhkan sebuah flowchart untuk memudahkan dalam menentukan alur perhitungan sebagai alur untuk menemukan hasil dari penerapan cluster terhadap data yang akan diproses.

Berikut adalah flowchart untuk menentukan clustering dengan hierarchichal clustering.

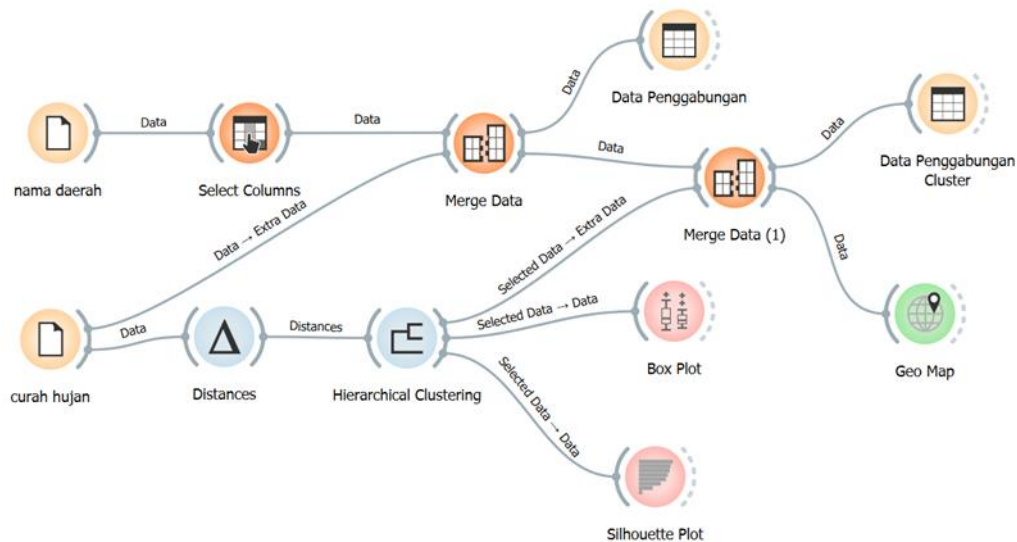


Gambar 1. Flowchart Analisis Clustering

Pada tahapan ini dilakukan analisis data banyaknya volume curah hujan di tiap provinsi yang ada di Indonesia. Data yang akan diperoleh kemudian akan diolah dengan menggunakan perhitungan bobot dari tiap indeks. Pada tahapan sebelumnya, telah ditentukan akan dicluster ke dalam 5 cluster.

HASIL DAN PEMBAHASAN

Clustering digunakan menggunakan tool orange. Melalui proses data mining dan metode hierarchical clustering.



Gambar 2. Alur Kerja Proses Klasifikasi Curah Hujan
Penggabungan dua dataset yaitu dari data nama provinsi dan curah hujan dilakukannya clustering dengan menggunakan marge data.

Tabel 3.
Dataset Longitude, Latitude, Dan Volume Hujan

Data Table					Sat Jan 06 24, 23:54:39				
Data instances: 33 Features: 3 Meta attributes: 1									
	name	longitude	latitude	volume_hujan					
1	Aceh	96.7494	4.69513	1575.0					
2	Sumatera Utara	99.5451	2.11535	975.9					
3	Sumatera Barat	100.8	-0.73994	3548.0					
4	Riau	101.707	0.293347	2048.3					
5	Jambi	102.438	-1.48518	1694.9					
6	Sumatera Selatan	103.914	-3.31944	1947.2					
7	Bengkulu	102.346	-3.57785	2668.9					
8	Lampung	105.407	-4.55858	1628.1					
9	Kepulauan Bangka Belitung	106.441	-2.74105	1534.7					
10	Kepulauan Riau	108.143	3.94565	2250.9					
11	DKI Jakarta	106.845	-6.21154	2169.5					
12	Jawa Barat	107.669	-7.09091	2199.3					
13	Jawa Tengah	110.14	-7.15097	1620.7					
14	DI Yogyakarta	110.426	-7.87538	2045.5					
15	Jawa Timur	112.238	-7.53606	2024.7					
16	Banten	106.064	-6.40582	1310.1					
17	Bali	115.189	-8.40952	1133.8					
18	Nusa Tenggara Barat	117.362	-8.65293	1147.9					
19	Nusa Tenggara Timur	121.079	-8.65738	1406.0					
20	Kalimantan Barat	111.475	-0.278781	2757.7					
21	Kalimantan Tengah	113.382	-1.68149	2748.4					
22	Kalimantan Selatan	115.284	-3.09264	2509.6					
23	Kalimantan Timur	116.419	1.64063	2069.4					
24	Sulawesi Utara	123.975	0.624693	1807.0					
25	Sulawesi Tengah	121.446	-1.43003	460.9					
26	Sulawesi Selatan	119.974	-3.6688	3382.0					
27	Sulawesi Tenggara	122.175	-4.14491	1589.6					
28	Gorontalo	122.447	0.699937	870.6					
29	Sulawesi Barat	119.232	-2.84414	1167.9					
30	Maluku	130.145	-3.23846	1987.2					
31	Maluku Utara	127.809	1.571	913.4					
32	Papua Barat	133.175	-1.33612	2844.6					
33	Papua	138.08	-4.26993	1265.9					

Penggabungan data kembali yang telah diclustering untuk mendapatkan cluster sesuai dengan tingkat kesamaan curah hujan di berbagai provinsi di Indonesia.

Tabel 4.
Data Cluster Yang Terbentuk Setiap Provinsi

Data Table

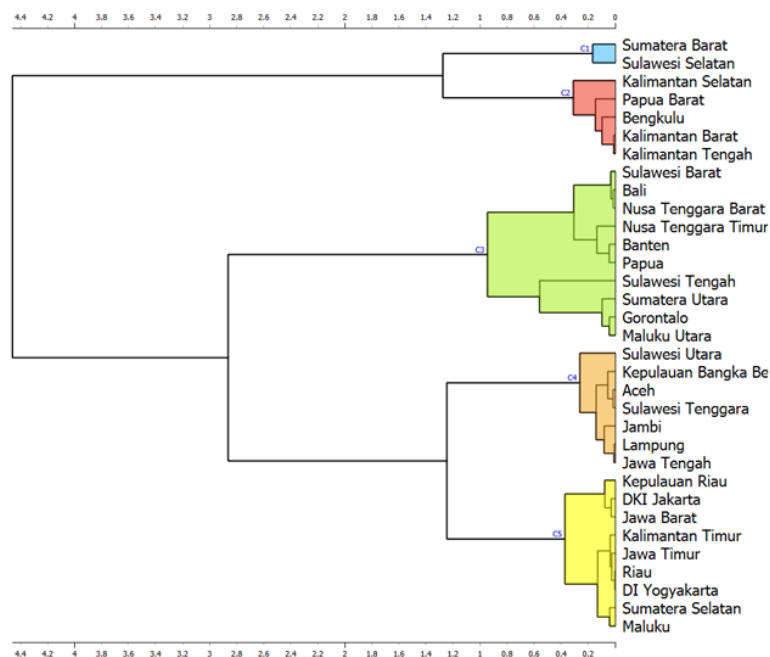
Sun Jan 07 24, 00:06:32

Data instances: 33
Features: 3
Meta attributes: 2

	name	Cluster	longitude	latitude	volume_hujan
1	Aceh	C4	96.7494	4.69513	1575.0
2	Sumatera Utara	C3	99.5451	2.11535	975.9
3	Sumatera Barat	C1	100.8	-0.73994	3548.0
4	Riau	C5	101.707	0.293347	2048.3
5	Jambi	C4	102.438	-1.48518	1694.9
6	Sumatera Selatan	C5	103.914	-3.31944	1947.2
7	Bengkulu	C2	102.346	-3.57785	2668.9
8	Lampung	C4	105.407	-4.55858	1628.1
9	Kepulauan Bangka Belitung	C4	106.441	-2.74105	1534.7
10	Kepulauan Riau	C5	108.143	3.94565	2250.9
11	DKI Jakarta	C5	106.845	-6.21154	2169.5
12	Jawa Barat	C5	107.669	-7.09091	2199.3
13	Jawa Tengah	C4	110.14	-7.15097	1620.7
14	DI Yogyakarta	C5	110.426	-7.87538	2045.5
15	Jawa Timur	C5	112.238	-7.53606	2024.7
16	Banten	C3	106.064	-6.40582	1310.1
17	Bali	C3	115.189	-8.40952	1133.8
18	Nusa Tenggara Barat	C3	117.362	-8.65293	1147.9
19	Nusa Tenggara Timur	C3	121.079	-8.65738	1406.0
20	Kalimantan Barat	C2	111.475	-0.278781	2757.7
21	Kalimantan Tengah	C2	113.382	-1.68149	2748.4
22	Kalimantan Selatan	C2	115.284	-3.09264	2509.6
23	Kalimantan Timur	C5	116.419	1.64063	2069.4
24	Sulawesi Utara	C4	123.975	0.624693	1807.0
25	Sulawesi Tengah	C3	121.446	-1.43003	460.9
26	Sulawesi Selatan	C1	119.974	-3.6688	3382.0
27	Sulawesi Tenggara	C4	122.175	-4.14491	1589.6
28	Gorontalo	C3	122.447	0.699937	870.6
29	Sulawesi Barat	C3	119.232	-2.84414	1167.9
30	Maluku	C5	130.145	-3.23846	1987.2
31	Maluku Utara	C3	127.809	1.571	913.4
32	Papua Barat	C2	133.175	-1.33612	2844.6
33	Papua	C3	138.08	-4.26993	1265.9

Hasil Hierarchical Clustering

Mesin melakukan clustering melalui Distances dengan metode Hierarchical Clustering. Hierarchical Clustering merupakan metode clustering untuk membagi data cluster menjadi hierarki. Hubungan hierarki antar cluster dapat dilihat menggunakan dendrogram dengan 5 cluster seperti gambar dibawah ini.

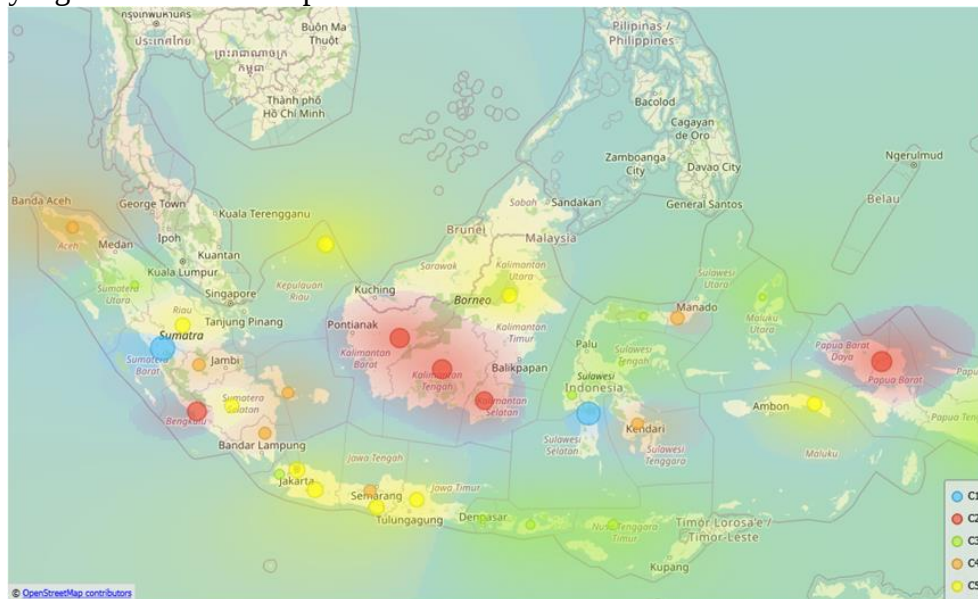


Gambar 3. Hierarchical Clustering

Setiap provinsi telah terclustering menjadi 5 cluster sesuai dengan karakteristik dari tiap cluster, yaitu C1(sumatera Barat dan Sulawesi Selatan), C2(Kalimantan Selatan, Papua Barat, Bengkulu, Kalimantan Barat, Kalimantan Tengah), C3(Sulawesi Barat, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Banten, Papua, Sulawesi Tengah, Sulawesi Utara, Gorontalo, Maluku Utara), C4(Sulawesi Utara, Kepulauan Bangka Belitung, Aceh, Sulawesi Tenggara, Jambi, Lampung, Jawa Tengah), C5(Kepulauan Riau, DKI Jakarta, Jawa Barat, Kalimantan Timur, Jawa Timur, Riau, DI Yogyakarta, Sumatera Selatan, Maluku).

Hasil Visualisasi Menggunakan Geomap

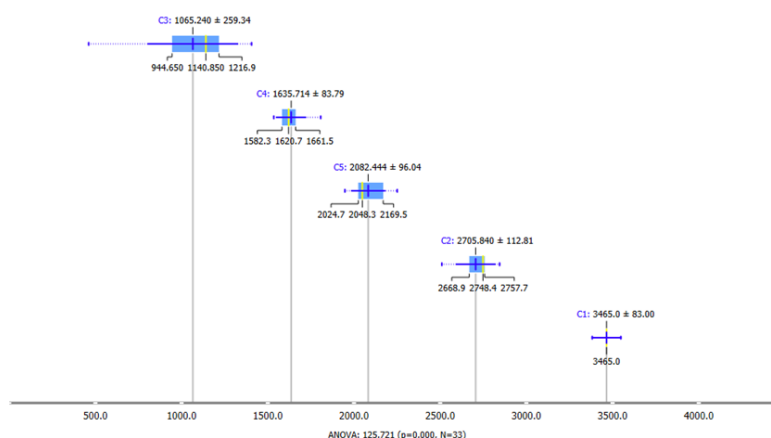
Hasil visualisasi menggunakan Geomap ditampilkan pada gambar 7. Terdapat lima cluster yang terlihat di Geomap.



Gambar 4. Hasil visualisasi menggunakan Geomap

Pada cluster 1 terdapat titik yang bewarna biru menandakan curah hujan sangat lebat, cluster 2 terdapat titik bewarna merah yang menandakan hujan lebat. Cluster 3 bewarna hijau yang menandakan hujan sedang, cluster 4 bewarna orange yang menandakan hujan ringan. Yang terakhir cluster 5 yang ditandai dengan titik bewarna kuning yang artinya tidak hujan.

Visualisasi Menggunakan Boxplot



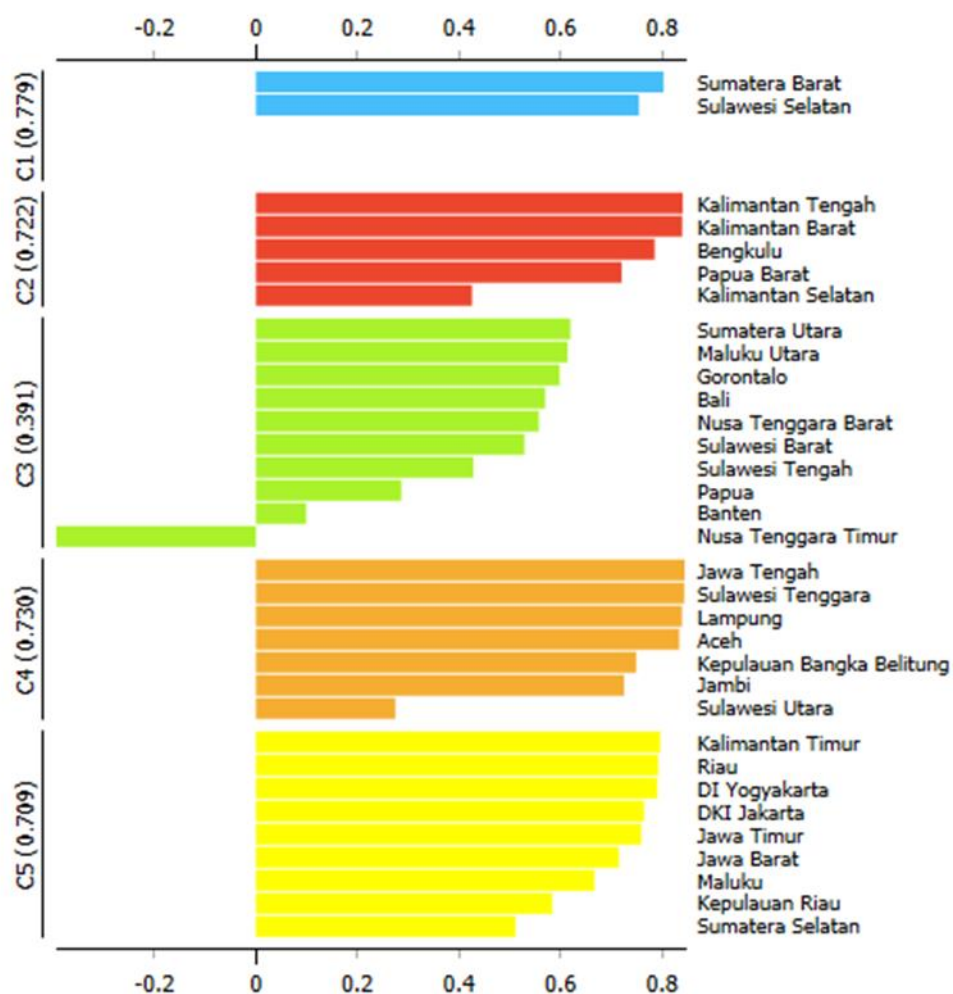
Gambar 5. Hasil visualisasi Boxpot

Pada Gambar diatas menunjukkan karakteristik dari curah hujan pada tiap cluster yang

memiliki range curah hujan yang berbeda-beda. Provinsi yang masuk ke dalam karakteristik cluster 3 memiliki paling banyak provinsi dengan curah hujan sedang. Sedangkan provinsi yang masuk ke dalam karakteristik cluster 1 memiliki paling sedikit provinsi dengan curah hujan sangat lebat.

Visualisasi Menggunakan Silhouette Plot

Data sudah memiliki nilai positif yang berarti data dari semua provinsi telah masuk kedalam cluster. Terdapat 2 provinsi yang masuk ke cluster 1 (Hujan sangat lebat) diantaranya provinsi Sumatra Barat dan Sulawesi Selatan. 5 provinsi yang masuk ke cluster 2 (Hujan lebat) antara lain Kalimantan Barat, Kalimantan Tengah, Bengkulu, Papua Barat, dan Kalimantan Selatan. Selanjutnya pada cluster 3 (Hujan sedang) terdapat 10 provinsi yang masuk kategori diantaranya Sumatra Utara, Maluku Utara, Gorontalo, Bali, Nusa Tenggara Barat, Sulawesi Barat, Sulawesi Tengah, Papua, Banten, Nusa Tenggara Timur. Pada cluster 4 (Hujan Ringan) terdapat 7 provinsi antara lain Jawa Tengah, Sulawesi Tenggara, Lampung, Aceh, Kepulauan Bangka Belitung, Jambi, Sulawesi Utara. Lalu ada 9 provinsi yang masuk kategori cluster 5 (Tidak Hujan) antara lain Kalimantan Timur, Riau, DI Yogyakarta, DKI Jakarta, Jawa Timur, Jawa Barat, Maluku, Kepulauan Riau, dan Sumatera Selatan.



Gambar 6. Hasil Visualisasi Shillhouette Plot

KESIMPULAN

Curah hujan pada tiap provinsi di Indonesia berbeda-beda. Terdapat beberapa faktor yang mempengaruhi kondisi cuaca. Antara lain kelembapan, suhu, angin, curah hujan, sinar matahari, dan tekanan atmosfer. Data mining dengan menggunakan hierarchical clustering yang direduksi menggunakan tool orange menghasilkan 5 cluster. Diantaranya Cluster 1 yaitu hujan sangat lebat, cluster 2 hujan lebat, cluster 3 adalah hujan sedang, cluster 4 yaitu hujan ringan, dan cluster 5 adalah tidak hujan. Klasifikasi curah hujan pada provinsi yang ada di Indonesia terdapat 2 provinsi yang masuk ke cluster 1, 5 provinsi yang masuk ke cluster 2, 10 provinsi yang masuk ke cluster 3, 7 provinsi yang masuk ke cluster 4 dan 9 provinsi yang masuk ke cluster 5. Klasifikasi ini juga di visualisasikan oleh Geomap, Bloxspot, shilluete untuk mempermudah proses klasifikasi.

DAFTAR PUSTAKA

- Agung, A. S., Fauzi, A. A., Risal, A. A. N., & Adiba, F. (2023). Implementasi Teknik Data Mining terhadap Klasifikasi Data Prediksi Curah Hujan BMKG Di Sulawesi Selatan. *Jurnal Tekno Insentif*, 17(1), 22-23.
- Ibrahim, M., & Goyal, S. (2023). Rainfall Classification Using Hierarchical Clustering Method: Lombok Case Study. *Journal of Tamora Science and Modeling*, 1(1), 38-45.
- Inayatus&Nabiilah AF. (2021). Introduction to Hierarchical Clustering, <https://algotech.netlify.app/blog/introduction-to-hierarchical-clustering/> artikel Algoritma Technical Blog.
- Jayadianti, H., Cahyadi, T. A., Amri, N. A., & Pitayandanu, M. F. (2020). Metode Komparasi Artificial Neural Network Pada Prediksi Curah Hujan-Literature Review. *Jurnal Tekno Insentif*, 14(2), 47-53.
- Juaeni, I. (2014). Dampak Penerapan Principal Component Analysis (Pca) Dalam Clustering Curah Hujan Di Pulau Jawa, Bali, Dan Lombok [Impact of Principal Component Analysis (Pca) Implementation on Rainfall Clustering Over Java, Bali and Lombok Islands]. *Jurnal Sains Dirgantara*, 11(2).
- Mulyanti, H. (2022). Identifikasi Wilayah Hujan Di Kabupaten Bojonegoro Menggunakan Teknik Klaster Aglomerasi. *Jurnal teknik sipil*, 7(2), 91-99.
- Mutoffar, M. M., Naseer, M., & Fadillah, A. (2022). KLASIFIKASI KUALITAS AIR SUMUR MENGGUNAKAN ALGORITMA RANDOM FOREST. *Naratif: Jurnal Nasional Riset, Aplikasi dan Teknik Informatika*, 4(2), 138-146.
- Mulyono, D. (2014). Analisis karakteristik curah hujan di wilayah Kabupaten Garut Selatan. *Jurnal Konstruksi*, 12(1).
- NAIBAH, L. P. N., & Windarto, A. P. (2023). Penerapan Data Mining Pada Tingkat Penghunian Kamar Pada Hotel Berbintang Berdasarkan Provinsi Menggunakan K-Means Clustering. *Jurnal Penelitian Ilmu Komputer*, 1(1), 32-41.
- Nuryaman, Y. (2018). Komparasi Algoritma Kmean dan Ahc untuk Klasifikasi Curah Hujan di Indonesia. *IKRA-ITH Informatika: Jurnal Komputer dan Informatika*, 2(2), 70-75
- Pratama, A. R. I., Latipah, S. A., & Sari, B. N. (2022). Optimasi klasifikasi curah hujan menggunakan support vector machine (svm) dan recursive feature elimination (rfe). *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 7(2), 314-324.
- Simanjuntak, K. P., & Khaira, U. (2021). Pengelompokan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering: Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 7-16.
- Suhirman, H. W., & Wintolo, H. (2019). System for Determining Public Health Level Using the Agglomerative Hierarchical Clustering Method. *Compiler*, 8(1), 95-104.
- Sundari, M. A., Pane, R., & Rohani, R. (2023). Data Mining Clustering Korban Kejahatan Pelecehan Seksual dengan Kekerasan Berdasarkan Provinsi Menggunakan Metode AHC. *Building of Informatics, Technology and Science (BITS)*, 5(1), 364-375.
- Susilowati, S., & Sadad, I. (2015). Analisa karakteristik curah hujan di Kota Bandar Lampung.

Konstruksia, 7(1).