

KOMPARASI ALGORITMA KMEAN DAN AHC UNTUK KLASIFIKASI CURAH HUJAN DI INDONESIA

Yosep Nuryaman¹, Ayuni Asistiyasari², Afri Yudha³

^{1,2,3}STMIK NUSA MANDIRI PASCA SARJANA

Jl. Damai No. 8, Warung Jati Barat, Pasar Minggu, Jakarta Selatan 12540

E-mail : yosepnuryaman@gmail.com¹, ayunioasis26@gmail.com², ibnugazali@gmail.com³

ABSTRAK

Curah hujan memiliki peranan penting di berbagai sektor seperti sektor pertanian, perkebunan hingga penerbangan. Dengan salah satu karakteristik negara beriklim tropis seperti di Indonesia yaitu curah hujan yang tinggi dan naik turun setiap tahunnya, perlu adanya pengelompokan curah hujan di wilayah Indonesia dengan harapan memberi acuan terhadap sektor-sektor penting tersebut.. K-Mean dan AHC adalah salah satu dari algoritma Clustering. Dengan membandingkan algoritma pengklusteran yaitu *Kmean* dan AHC, maka didapatkan hasil pengkluster terbaik k3 pada literasi ke 4 dengan rasio 0,000249.

Kata kunci : Curah hujan, Data mining, Klustering, *kmean*, AHC,

1. PENDAHULUAN

Indonesia sebagai negara beriklim tropis memiliki 2 musim yaitu musim hujan dan musim kemarau. Karakteristik negara beriklim tropis juga memiliki curah hujan yang tinggi. Namun dengan adanya pemanasan global, keadaan curah hujan di berbagai wilayah Indonesia naik turun setiap tahunnya. Padahal curah hujan memiliki peranan penting terhadap berbagai sektor seperti perkebunan, pertanian hingga penerbangan.

Curah hujan berpengaruh nyata dalam meningkatkan produksi TBS sedangkan hari hujan berpengaruh nyata terhadap penurunan produksi TBS pada tanaman kelapa sawit berumur 5 tahun di kebun Begerpang Estate PT.PP London Sumatra Indonesia, Tbk [8].

Metode K-Means telah diterapkan untuk mengelompokkan curah hujan ke dalam tiga kategori (rendah, sedang dan tinggi) di Propinsi Kalimantan Timur. Hasil penelitian menunjukkan bahwa metode K-Means clustering dengan tiga cluster merupakan jumlah cluster yang ideal dibandingkan dengan hasil clustering berjumlah dua cluster [4].

Pengujian dilakukan dengan membandingkan hasil clustering kedua algoritma tersebut dengan pengelompokan berdasarkan aturan BAPPEDA 2008. Hasil clustering menggunakan algoritma FCM memiliki tingkat akurasi hanya 50%, sedangkan

untuk algoritma K-means memiliki tingkat akurasi lebih baik yaitu 83.33% [7].

Atas dasar hal tersebut, penulis mengangkat tema komparasi algoritma kmean dan AHC untuk klasifikasi curah hujan di Indonesia. Dengan adanya penelitian ini diharapkan mampu memberikan klasifikasi curah hujan terbaik di Indonesia yang dapat digunakan sebagai data pendukung dalam berbagai sektor..

2. METODOLOGI

Penulis dalam mengumpulkan data-data pendukung melakukan kegiatan observasi, wawancara dan studi pustaka dan menghasilkan data seperti yang tampak dibawah ini :

Tabel II.1
Data Penelitian

Tahun	2013	2014	2015
Provinsi			
Aceh	1,623	2,264	1,575
Sumatera Utara	2,627	2,148	975
Sumatera Barat	4,627	2,838	3,548
Riau	2,628	2,343	2,048
Jambi	2,093	1,781	1,694
Sumatera Selatan	3,409	1,668	1,947

Sumber : BPS (2017) [2]

Dalam penelitian ini menggunakan model CRISP-DM (Cross-Industry Standard Proses for Data Mining), yang terdiri dari 6 tahap yaitu :

1. Business/Research Understanding Phase
2. Data Understanding Phase
3. Data Preparation Phase
4. Modeling Phase
5. Evaluation Phase
6. Deployment Phase

3. LANDASAN TEORI

Data Mining

Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran computer (machine learning) untuk menganalisis dan mengekstarksi pengetahuan (knowledge) secara otomatis [1].

Menambahkan definisi lain dari data mining merupakan proses iterative dan interaktif untuk menemukan pola atau model baru yang sah (Sempurna), bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive database [1].

KDD (Knowledge discovery in database)

Istilah data mining dan KDD seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar [6]. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu hapan dalam keseluruhan proses KDD adalah data mining.

Klustering

Analisis kluster atau clustering merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok lain[5]..

K-MEAN

K-means clustering termasuk dalam partitioning clustering yang disebut juga exclusive clustering yang memisahkan data ke k daerah bagian yang terpisah dan setiap data harus termasuk ke dalam cluster tertentu dan memungkinkan setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke cluster yang lain. K-means merupakan algoritma yang sangat terkenal karena kemudahan dan kemampuannya untuk melakukan pengelompokan data besar dengan sangat cepat [3].

Algoritma k-means clustering :

[1] Inisialisasi clustering centroid $\mu_1, \mu_2, \dots, \mu_k, \in \mathbb{R}^n$

[2] Repeat until convergence :

$$\left\{ \begin{array}{l} \text{For } i, c^{(i)} := \arg \min_j \|x^{(i)} - \mu_j\|^2 \quad (1) \end{array} \right.$$

$$\text{For } j, \mu_j := \frac{\sum_{i=1}^m 1_{\{c^{(i)}=j\}} x^{(i)}}{\sum_{i=1}^m 1_{\{c^{(i)}=j\}}}$$

AHC

Hierarchical clustering adalah metode clustering yang mengelompokkan data dengan urutan partisi berkalang, metode ini dikelompokkan menjadi dua metode yaitu agglomerative dan divisive, metode agglomerative berawal dari obyek-obyek individual dimana pada awalnya banyaknya cluster sama dengan banyaknya obyek [9].

4. HASIL DAN PEMBAHASAN

Dan berikut hasil perhitungan menggunakan algoritma K-mean dan AHC

Hasil Perhitungan K-mean

Proses awal perhitungan Kmean adalah menentukan centroid dengan cara random berdasarkan data training yang ada seperti yang tampak pada tabel dibawah ini

Table VI.1
Centroid Literasi 1

K3	X	Y	Z
C1	2309	2025	2045
C2	2456	1682	1628
C3	4627	2838	3548

Sumber : Pengolahan data (2017)

Kemudian dengan rumus yang ada, cari keanggotaan kluster dari setiap kota. Dan berikut hasil keanggotaan berdasarkan literasi 1.

Table VI.2
Keanggotaan Literasi 1

Data	Jarak ke Centroid			Keanggotaan
	C1	C2	C3	
1	865	1018	3639	C1
2	1123	819	3331	C2
3	2880	3120	0	C3
4	451	802	2547	C1
5	479	382	3312	C2
6	1161	1005	2327	C2
7	2206	2469	1194	C3
8	560	0	3120	C2
9	816	394	2933	C2
10	1512	1779	1807	C1
11	916	1341	2515	C1
12	542	935	2410	C1
13	803	961	2785	C1
14	0	560	2880	C1
15	1547	1172	2803	C2
16	1002	581	3657	C2
17	1031	610	3712	C2
18	898	464	3571	C2
19	1443	1809	1585	C1
20	906	1246	1981	C1
21	674	949	2346	C1
22	1644	1719	1963	C1
23	2250	2548	682	C3
24	1432	1056	4167	C2
25	1423	1075	4168	C2
26	1584	1926	1398	C3
27	2019	1928	2361	C2

Sumber : Pengolahan data (2017)

Dan dari tabel tersebut didapatkan rasio dari literasi 1 dengan nilai 1,000170. Dikarenakan ini masih literasi 1 maka literasi dilanjutkan. Sama seperti cara sebelumnya, langkah selanjutnya adalah menentukan centroid terbaru. Dan berikut adalah hasilnya

Table VI.3
Centroid Literasi 2

	X	Y	Z
C1	2784	2555	2095
C2	2574	1694	1340
C3	4000	2931	3111

Sumber : Pengolahan data (2017)

Seperti pada langkah sebelumnya maka berikut hasil keanggotaan literasi 2

Table VI.4
Keanggotaan Literasi 2

Data	Jarak ke Centroid			Keanggotaan
	C1	C2	C3	
1	1304	1133	2907	C2
2	1201	585	2657	C2
3	2364	3225	770	C3
4	266	962	1832	C1
5	1112	604	2639	C2
6	1095	1033	1816	C2
7	1534	2529	591	C3
8	1042	311	2478	C2
9	1045	329	2326	C2
10	806	1836	1063	C1
11	444	1471	1750	C1
12	222	1110	1692	C1
13	505	977	2048	C1
14	713	823	2194	C1
15	1519	1014	2327	C2
16	1468	470	2996	C2
17	1532	529	3056	C2
18	1475	513	2938	C2
19	862	1949	828	C3
20	505	1418	1290	C1
21	152	1068	1630	C1
22	1018	1683	1337	C1
23	1762	2686	333	C3
24	1960	971	3507	C2
25	2049	1087	3537	C2
26	1019	2063	648	C3
27	1509	1791	1856	C1

Sumber : Pengolahan data (2017)

Dan rasio yang didapatkan pada literasi ke 2 adalah 0,000223. Karena rasio sekarang lebih besar dari rasio sebelumnya maka literasi dilanjutkan hingga rasio tidak lebih besar ataupun keanggotaannya tidak berubah.

Selanjutnya dilakukan penentuan centroid baru seperti yang tampak dibawah ini

Table VI.5
Centroid Literasi 3

	X	Y	Z
C1	2977	2572	1999
C2	2374	1655	1366
C3	3852	2916	3038

Sumber : Pengolahan data (2017)

Dan berikut hasil keanggotaan literasi 3.

Table VI.6
Keanggotaan Literasi 3

Data	Jarak ke Centroid			Keanggotaan
	C1	C2	C3	
1	1451	989	2744	C2
2	1161	678	2519	C2
3	2279	3353	931	C3
4	420	1002	1674	C1
5	1224	450	2487	C2
6	1003	1188	1715	C1
7	1421	2658	565	C3
8	1095	276	2336	C2
9	1019	496	2197	C2
10	690	1949	925	C1
11	589	1496	1586	C1
12	402	1152	1533	C1
13	518	1037	1895	C1
14	864	777	2039	C2
15	1390	1208	2238	C2
16	1514	319	2852	C2
17	1585	362	2911	C2
18	1541	331	2796	C2
19	849	2033	663	C3
20	550	1490	1137	C1
21	206	1146	1476	C1
22	811	1844	1241	C1
23	1713	2792	405	C3
24	2020	816	3361	C2
25	2131	910	3393	C2
26	987	2156	483	C3
27	1295	1980	1791	C1

Sumber : Pengolahan data (2017)

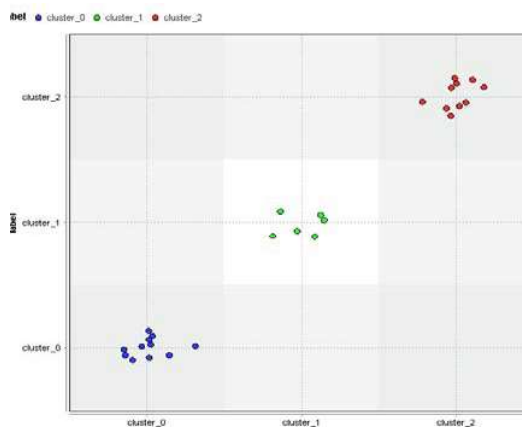
Perhitungan pada literasi ke 4 dengan nilai rasio 0,000249 menjadi literasi terakhir dikarenakan keanggotaannya tidak berubah. Dan berikut keanggotaan pada literasi ke 4.

Table VI.7
Keanggotaan Literasi 4

Data	Jarak ke Centroid			Keanggotaan
	C1	C2	C3	
1	1545	900	2744	C2
2	1178	702	2519	C2
3	2212	3400	931	C3
4	501	1004	1674	C1
5	1282	384	2487	C2
6	926	1265	1715	C1
7	1371	2692	565	C3
8	1121	308	2336	C2
9	1005	580	2197	C2
10	662	1974	925	C1
11	699	1479	1586	C1
12	480	1156	1533	C1
13	596	1034	1895	C1
14	932	753	2039	C2
15	1314	1303	2238	C2
16	1550	275	2852	C2
17	1621	315	2911	C2
18	1573	302	2796	C2
19	841	2052	663	C3
20	552	1511	1137	C1
21	272	1164	1476	C1
22	723	1891	1241	C1
23	1663	2829	405	C3
24	2062	768	3361	C2
25	2172	865	3393	C2
26	962	2180	483	C3
27	1206	2043	1791	C1

Sumber : Pengolahan data (2017)

Selain itu penulis juga mencoba menampilkan penyebaran menggunakan software rapidminer. Dan berikut hasil penyebaran hasil pemrosesan menggunakan rapidminer.



Sumber : Pengolahan data (2017)

Gambar 1
Penyebaran Kota

Hasil Perhitungan AHC

Dengan rumus yang ada didapatkan hasil matrix literasi 1 dengan hasil kedekatan terdekat antara data ke 17 dan data ke 16 dengan nilai 147 seperti yang tampak pada tabel dibawah ini

Table VI.7
Matrix Literasi 1

	1	2	16	17	26	27
1	0	1,719	1,597	1,603	3,625	3,185
2	1,719	0	1,137	1,284	3,337	2,279
16	1,597	1,137	0	147	4,159	3,100
26	3,625	3,337	4,159	4,278	0	2,286
27	3,185	2,279	3,100	3,219	2,286	0

Sumber : Pengolahan data (2017)

Kemudian dari matrix literasi 2 didapatkan kedekatan data terkecil antara data ke 4 dan ke 12 yaitu sebesar 249 seperti yang tampak pada tabel dibawah ini.

Table VI.8
Matrix Literasi 2

Data	1	2	3	4	12	26	27
1	0	1,719	5,551	1,558	1,806	3,625	3,185
2	1,719	0	5,263	1,270	1,518	3,337	2,279
3	5,551	5,263	0	3,993	3,745	1,926	2,984
4	1,558	1,270	3,993	0	249	2,068	2,574
12	1,806	1,518	3,745	249	0	1,819	2,627
26	3,625	3,337	1,926	2,068	1,819	0	2,286
27	3,185	2,279	2,984	2,574	2,627	2,286	0

Sumber : Pengolahan data (2017)

Langkah tersebut diulangi terus. Untuk Literasi ke 18 akan tampak seperti tabel dibawah ini.

Table VI.9
Matrix Literasi 18

	(4,12:2 1:20)	(16,17:18: 5:24,25)	(19,26: 10:7)	(11,1 3:14)	1
(4,12:21:2 0)	0	3940	2952	2617	21
(16,17:18: 5:24,25)	3940	0	6026	3655	17
(19,26:10: 7)	2952	6026	0	3592	10
(11,13:14)	2617	3655	3592	0	21
1	2136	1716	4510	2139	0

Sumber : Pengolahan data (2017)

Dan berikut hasil dari literasi ke 22 dengan jarak terdekat 2761.

Table VI.10
Matrix Literasi 22

	(3 ,2 3)	(19,2 6:10: 7)	(16,17:1 8:5:24,2 5:1)	(8,9:2: 11,13: 14)	(4,12:21:2 0:6,15:22, 27))
(3,23)	0	2761	7067	5263	4610
(19,26:10: 7)	27 61	0	6026	4222	3569
(16,17:18: 5:24,25:1)	70 67	6026	0	3655	4415
(8,9:2:11, 13:14)	52 63	4222	3655	0	3295
(4,12:21:2 0:6,15:22, 27))	46 10	3569	4415	3295	0

Sumber : Pengolahan data (2017)

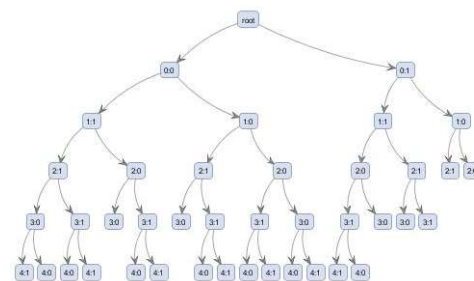
Langkah tersebut diulangi hingga berhenti pada literasi ke 26 seperti yang tampak pada tabel dibawah ini.

Table VI.11
Matrix Literasi 26

Data	(19,26: 10:7:3, 23)	(8,9:2:11,13:14:4,12:21: 20:6,15:22,27:16,17:18: 5:24,25:1)
(19,26:10:7:3,23)	0	7066.9
(8,9:2:11,13:14:4,12:21: 20:6,15:22,27:16,17:18: 5:24,25:1)		0

Sumber : Pengolahan data (2017)

Dan berikut hasil pengolahan Algoritma AHC menggunakan rapidminer



Sumber : Pengolahan data (2017)

Gambar 2
Tree Clustering AHC

Perbandingan Hasil Perhitungan Kmean dan AHC

Berdasarkan Perhitungan sebelumnya berikut perbandingan keanggotaan k-mean dan AHC

Table VI.12
Matrix Literasi 1

No	Nama Kota	Cluster Berdasarkan	
		Kmean	AHC
1	Aceh	C2	C2
2	Sumatera Utara	C2	C1
3	Sumatera Barat	C3	C3
4	Riau	C1	C1
5	Jambi	C2	C2
6	Sumatera Selatan	C1	C1
7	Bengkulu	C3	C3
8	Lampung	C2	C1
9	Kepulauan Bangka Belitung	C2	C1
10	Kepulauan Riau	C1	C3
11	DKI Jakarta	C1	C1
12	Jawa Barat	C1	C1
13	Jawa Tengah	C1	C1
14	DI Yogyakarta	C2	C1
15	Banten	C2	C1
16	Bali	C2	C2
17	Nusa Tenggara Barat	C2	C2
18	Nusa Tenggara Timur	C2	C2
19	Kalimantan Tengah	C3	C3
20	Kalimantan Selatan	C1	C1
21	Kalimantan Timur	C1	C1
22	Sulawesi Utara	C1	C1
23	Sulawesi Selatan	C3	C3
24	Gorontalo	C2	C2
25	Sulawesi Barat	C2	C2
26	Papua Barat	C3	C3
27	Papua	C1	C1

Sumber : Pengolahan data (2017)

5. KESIMPULAN

Dari hasil perhitungan yang ada penulis menarik kesimpulan

- Anggota C1 KMean adalah 4,6,10,11,12,13,20,21,22,27
- Anggota C1 AHC adalah 2,4,6,8,9,11,12,13,14,15,20,21,22,27
- Anggota C2 K mean adalah 1,2,5,8,9,14,15,16,17,18,24,25
- Anggota C2 AHC adalah 1,5,16,17,18,24,25

- Anggota C3 Kmean adalah 3,7,19,23,26
- Anggota C3 AHC adalah 3,7,10,19,23,26

dan algoritma kmean lebih baik dalam menentukan keanggotaan dengan nilai rasio 0,000249. Selain itu di setiap literasi adanya peregeseran titik tengah dengan nilai centroid sehingga hasil literasinya juga lebih cepat..

DAFTAR PUSTAKA

- Aprilia, Dennis, Donny Aji Baskoro, Lia Ambarwati dan I Wayan Simri Wicaksana. 2013. Belajar Data Mining dengan Rapid Miner. Jakarta : Gramedia Pustaka Utama.
- Badan Pusat Statistik.(2017). (<https://www.bps.go.id/index.php/linkTabelStatis/1959>), diakses 20 April 2017.
- Hartigan, J.A., & Wong, M.A. (1979). *Algorithm AS 136 : A K-means clustering algorithm*. Applied statistics, 100-108.
- Hasanah, Nur, Muh. Ugiarto, Novianti Puspitasari (2017). Sistem Pengelompokan Curah Hujan Menggunakan K-Means di Wilayah Kalimantan Timur. Prosiding Seminar Nasional Ilmu Komputer dan Teknologi Informasi. Vol.2. No.2. 122-126.
- Jang, J.S.R., Sun, C.T., dan Mizutani, E (2004). *Neuro-Fuzzy AND Soft Computing*. Singapura : Pearson Education
- Kusrini dan Emha Taufiq Luthfi (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset
- Ulfah, Aniq, Noviciantie, Shofwatul Uyun (2015). Analisis Kinerja Algoritma Fuzzy C-Means dan K-Means Pada Data Kemiskinan. Jatsi. Vol 1 No.2.139-147
- Ningsih Lastiar Simanjuntak, Rosita Sipayung, Irsal (2014). Pengaruh Curah Hujan dan Hari Hujan Terhadap Produksi Kelapa Sawit Berumur 5,10 Dan 15 Thun di Kebun Begerpang Estate PT. PP London Sumatra Indonesia, Tbk. Jurnal Online Agroekoteknologi. Vol.2 No.3 : 1141-1151.
- Xu, Rui, C. Wunsch (2009). Clustering, John Wiley & Sons, INC.