

PENERAPAN ALGORITMA *K-MEANS* DALAM PENGELOMPOKAN CURAH HUJAN DI DAERAH JABODETABEK

Yoga Religia, Ari Susanto

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa
yoga@pelitabangsa.ac.id

Disetujui, 05 Maret 2018

Abstrak

Akhir-akhir ini cuaca semakin sulit untuk diprediksi, bahkan terkadang terjadi hujan saat musim kemarau dan sebaliknya. Hal ini sangat berpengaruh pada aktifitas sehari-hari seperti keselamatan masyarakat, sosial ekonomi, produksi pertanian, perkebunan, perikanan, penerbangan, dan sebagainya dalam suatu daerah. Maka dari itu informasi yang akurat tentang kondisi cuaca sangatlah penting sehingga kita dapat mempersiapkan diri. Dalam penyampaian informasi agar lebih akurat perlu adanya penelitian tentang cuaca atau curah hujan. Dalam penelitian ini kita akan mengambil contoh kondisi cuaca di wilayah Jakarta dan sekitarnya dimana wilayah tersebut lebih beragam aktivitasnya. Algoritma *K-Means* adalah metode yang dapat digunakan dalam penelitian ini, yaitu untuk melakukan pengelompokan curah hujan. Data yang diambil adalah data dari BMKG untuk wilayah Jakarta dan sekitarnya dimana nantinya akan dikelompokkan menjadi dua kelompok yaitu curah hujan tinggi dan rendah menggunakan algoritma *K-Means* dan metode perhitungan jarak Euclidean Distance. Pengelompokan curah hujan ini nantinya dapat memberikan informasi yang akurat sehingga dapat mengurangi dampak perubahan kondisi cuaca yang secara mendadak.

Kata Kunci : Algoritma *K-Means*, Euclidean Distance, Pengelompokan Cuaca, Data Mining

ABSTRACT

*Lately the weather is increasingly difficult to predict, even sometimes there is rain during the dry season and vice versa. This is very influential on daily activities such as community safety, socio-economic, agricultural production, plantations, fisheries, aviation, and so on in an area . Therefore accurate information about weather conditions is very important so we can prepare ourselves. In delivering information to be more accurate there needs to be research on weather or rainfall. In this study we will take an example of weather conditions in the Jakarta and around region where the region has more diverse activities. The *K-Means* algorithm is a method that can be used in this study, namely to group rainfall. The data taken is data from BMKG for the Jakarta and around region where later it will be grouped into two groups, high and low rainfall using the *K-Means* algorithm and Euclidean Distance distance calculation method. This clustering of rainfall can later provide accurate information so that it can reduce the impact of sudden changes in weather conditions*

Keywords: K-Means Algorithm, Euclidean Distance, Weather Clustering, Data Mining

1. Pendahuluan

Teknologi dan informasi berkembang dengan pesat sampai saat ini dan hampir menyentuh banyak bidang dan mempermudah kita dalam kegiatan sehari-hari yang kita lakukan. Salah satu teknologi tersebut yaitu komputer. Seiring dengan perkembangan jaman, peran komputer semakin banyak didalam kehidupan masyarakat (Rismawan & Kusumadewi, 2008). Dalam penerapannya, komputer dirancang untuk mempermudah atau meringankan pekerjaan manusia (Nugroho & Fatta, 2012), salah satunya yaitu untuk mengolah data dalam jumlah besar yang mana akan sulit membaca dan mengetahui pola serta relasi data jika dilakukan secara manual (Suyanto, 2017).

Posisi wilayah Indonesia yang berada digaris katulistiwa membuat indonesia beriklim tropis, yang artinya memiliki dua musim yaitu musim kemarau dan musim penghujan (Athoillah, Mariana, & D, 2017). Hal itu secara tidak langsung berpengaruh terhadap aktifitas masyarakat. Akan tetapi akhir-akhir ini cuaca semakin sulit untuk diprediksi, bahkan terkadang terjadi hujan saat musim kemarau dan sebaliknya (Mutia, Sundoro, & Yajidin, 2017). Oleh karena itu, informasi cuaca yang akurat sangatlah penting karena besar

pengaruhnya terhadap segala macam aktifitas kehidupan seperti keselamatan masyarakat, sosial ekonomi, produksi pertanian,

perkebunan, perikanan, penerbangan, dan sebagainya dalam suatu daerah (Puspitasari & Havilludin, 2016). Dalam mendapatkan informasi yang akurat mengenai cuaca, terdapat beberapa sumber yang dapat dijadikan sebagai referensi dalam pengambilan data, salah satunya adalah data dari Badan Meteorologi Klimatologi dan Geofisika (BMKG). Kemudian dari data cuaca tersebut perlu adanya pengolahan informasi tentang cuaca yang baik dengan memanfaatkan metode yang tepat untuk mengelompokkan dua kategori cuaca yaitu curah hujan tinggi dan rendah (Puspitasari & Havilludin, 2016). Metode yang dapat dipergunakan dalam pengolahan data salah satunya menggunakan data mining (Chandra, 2016).

Secara umum, kegunaan data mining dapat di bagi menjadi dua yaitu, deskriptif dan, prediktif. Tujuan dari tugas prediktif adalah untuk memprediksi nilai dari atribut tertentu berdasarkan pada nilai dari atribut-atribut lain. Atribut yang diprediksi umumnya dikenal sebagai target atau variabel tak bebas, sedangkan atribut-atribut yang digunakan untuk membuat prediksi dikenal sebagai explanatory atau variabel bebas. Yang kedua yaitu deskriptif, tujuan dari tugas deskriptif adalah untuk menurunkan polapola (korelasi, trend, cluster, trayektori, dan anomali) yang meringkas hubungan yang pokok dalam data. Tugas data mining deskriptif sering merupakan penyelidikan dan seringkali memerlukan teknik postprocessing untuk validasi dan penjelasan hasil. Sedangkan secara fungsionalnya data mining bisa dikelompokkan menjadi enam yaitu klasifikasi, klasterisasi, regresi, deteksi anomali, asosiasi, dan perangkuman (Suyanto, 2017). Clustering sendiri merupakan sebuah algoritma yang dapat mengelompokkan data berdasarkan nilai-nilai patokan yang diberikan dan mengkalkulasikan kelompok data tersebut (Nasution & Eka, 2018). Salah satu algoritma clustering yang populer saat ini adalah K-Means (Nasution M. , 2017) . Algoritma K-Means terbukti cukup andal dan banyak digunakan dalam komunitas peneliti data mining dan berbagai macam aplikasi kecil hingga menengah karena kemudahan implementasinya (Suyanto, 2017). K-Means merupakan salah satu metode data clustering non hirarki yang digunakan untuk mempartisi data yang ada kedalam bentuk dua atau lebih cluster/kelompok. Metode K-Means mempartisi data kedalam cluster sehingga data yang memiliki karakteristik sama dikelompokkan kedalam satu cluster yang sama (Rismawan & Kusumadewi, 2008). Dalam melakukan pengelompokan algoritma K-Means membutuhkan perhitungan jarak untuk menghitung jarak terdekat antara suatu instance data ke sebuah titik centroid (Dutt, Ismail, & Herawan, 2017). Metode perhitungan jarak Euclidean Distance banyak digunakan dikarenakan memiliki akurasi yang cukup baik (Urip, Adi, & Widodo, 2017). Setelah perhitungan jarak dilakukan proses selanjutnya adalah melakukan pengujian terhadap data tersebut menggunakan Davies Bouldin Index (DBI). Pengukuran menggunakan Davies Bouldin Index bertujuan untuk memaksimalkan jarak antar cluster. Dalam penelitian ini, DBI digunakan untuk melakukan validasi data pada setiap cluster (Mustofa & Suasana, 2018).

2. Tinjauan Studi

Teori Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Ridwan , Suryono, & Sarosa, 2013). Dalam membangun sebuah data mining diperlukan kumpulan data training yang menjadi dasar acuan prediksi, kemudian metode classification ditentukan sesuai dengan karakteristik data training (Maesaroh & Kusriani, 2017). Secara garis besar metode pelatihan yang digunakan dalam teknik- teknik data mining dibedakan ke dalam dua pendekatan, yaitu:

1. Supervised learning, yaitu metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini, untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai output atau label selama proses training (Ridwan , Suryono, & Sarosa, 2013).
2. Unsupervised learning adalah teknik machine learning yang menggunakan prosedur yang berusaha mencari partisi dari sebuah pola. Unsupervised learning mempelajari bagaimana sistem dapat belajar untuk merepresentasikan pola input dengan cara menggambarkan struktur statistik dari keseluruhan input. Metode ini tidak memiliki target output yang eksplisit atau tidak ada pengklasifikasian input (Akbari, Novianty, & Setianingsih, 2017).

Ada beberapa teknik yang dimiliki data mining berdasarkan tugas yang bisa dilakukan, yaitu (Ridwan , Suryono, & Sarosa, 2013):

1. Description (deskripsi), Para peneliti biasanya mencoba menemukan cara untuk mendeskripsikan pola dan trend yang tersembunyi dalam data.

2. Estimation (estimasi), Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih kearah numerik dari pada kategori.
3. Prediction (prediksi), Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi dimasa depan).
4. Clasification (klasifikasi), Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, kita akan mengklasifikasikan pendapatan dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.
5. Clustering (pengelompokan), Clustering lebih ke arah pengelompokan record, pengamatan, atau kasus dalam kelas yang memiliki kemiripan.
6. Association (asosiasi), Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

Tahapan Data Mining

Terdapat beberapa tahap yang harus dilewati sebelum menambang data, antara lain yaitu (Huda, 2010):

1. Pembersihan data (data cleaning)
Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performasi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.
2. Integrasi data (data integration)
Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru.
3. Seleksi data (data selection)
Data yang ada pada database sering kali tidak semua dipakai, hanya data yang sesuai untuk dianalisis yang diambil dari database (Novianti & Aziz, 2015). Karena tidak semua tabel digunakan maka perlu dilakukan pembersihan data agar data yang akan diolah benar-benar relevan dengan yang dibutuhkan.
4. Transformasi data (Data Transformation)
Data digabung ke dalam format yang sesuai untuk diproses dalam data mining. Transformasi data merupakan proses perubahan atau penggabungan data ke dalam format yang sesuai.
5. Proses mining
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. Pengujian dan Evaluasi Sistem
Pengujian dilakukan dengan training dan testing menggunakan metode Cross validation (Novianti & Aziz, 2015).

Aturan Clustering

Menurut Han dan Kamber, 2012, syarat sekaligus tantangan yang harus dipenuhi oleh suatu algoritma clustering adalah:

1. Skalabilitas
Suatu metode clustering harus mampu menangani data dalam jumlah yang besar. Saat ini data dalam jumlah besar sudah sangat umum digunakan dalam berbagai bidang misalnya saja suatu database. Tidak hanya berisi ratusan objek, suatu database dengan ukuran besar bahkan berisi lebih dari jutaan objek.
2. Kemampuan analisa beragam bentuk data
Algoritma klasterisasi harus mampu dimplementasikan pada berbagai macam bentuk data seperti data nominal, ordinal maupun gabungannya.
3. Menemukan cluster dengan bentuk yang tidak terduga
Banyak algoritma clustering yang menggunakan metode Euclidean atau Manhattan yang hasilnya berbentuk bulat. Padahal hasil clustering dapat berbentuk aneh dan tidak sama antara satu dengan yang lain. Karenanya dibutuhkan kemampuan untuk menganalisa cluster dengan bentuk apapun pada suatu algoritma clustering.
4. Kemampuan untuk dapat menangani noise
Data tidak selalu dalam keadaan baik. Ada kalanya terdapat data yang rusak, tidak dimengerti atau hilang. Karena system inilah, suatu algoritma clustering dituntut untuk mampu menangani data yang rusak.

5. Sensitifitas terhadap perubahan input
Perubahan atau penambahan data pada input dapat menyebabkan terjadi perubahan pada cluster yang telah ada bahkan bisa menyebabkan perubahan yang mencolok apabila menggunakan algoritma clustering yang memiliki tingkat sensitifitas rendah.
6. Mampu melakukan clustering untuk data dimensi tinggi
Suatu kelompok data dapat berisi banyak dimensi ataupun atribut. Untuk itu diperlukan algoritma clustering yang mampu menangani data dengan dimensi yang jumlahnya tidak sedikit.

Interpresasi dan kegunaan Hasil dari clustering harus dapat diinterpretasikan dan berguna.

Algoritma K-Means

K-Means adalah salah satu algoritma pengelompokan yang menggunakan metode partisi. K-Means adalah algoritma pengelompokan yang membagi setiap item data kedalam sebuah cluster (Slamet, Rahman, Ramdhani, & Darmalaksana, 2016). Algoritma K-Means terdiri dari dua fase terpisah. Pada tahap pertama menghitung centroid k dan pada fase kedua menghitung setiap titik cluster yang terdekat centroid dari titik data masing-masing. Ada berbagai metode untuk menentukan jarak pusat terdekat dan salah satu metode yang paling banyak digunakan adalah euclidean distance (Dhanachandra, Manglem, & Chanu, 2015).

Metode K-Means sangat terkenal karena kemudahannya dan kemampuannya untuk mengelompokkan data besar dan outlier dengan sangat cepat. Dalam metode K-Means setiap data harus termasuk kecluster tertentu pada suatu tahapan proses, pada

$$\sum /$$

.....(2.1)

tahapan proses berikutnya dapat berpindah kecluster yang lain (Abdillah, Putra, & Renaldi, 2016). Berikut ini adalah langkah-langkah dalam algoritma K-Means (Puspitasari & Havilludin, 2016):

1. Tentukan k sebagai jumlah cluster yang akan dibentuk,
2. Tentukan k centroid awal secara random atau acak, dengan rumus sebagai berikut:
Pada persamaan 2.1, dimana, v adalah centroid pada cluster; adalah objek ke-i; n adalah banyaknya objek/jumlah objek yang menjadi anggota cluster.
3. Hitung jarak setiap objek ke masing-masing centroid dari masing-masing cluster dengan menggunakan metode Euclidian Distance, dengan rumus sebagai berikut :

$$\sqrt{\sum}$$

.....(2.2)

Pada persamaan 2.2, dimana, adalah jarak antara data ke-i dengan centroid ke-k; m adalah jumlah atribut; adalah data ke-i; adalah data pusat klaster ke-k.

4. Alokasikan masing-masing objek ke dalam centroid yang paling dekat,
5. Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan (2.1).
6. Ulangi langkah 3 jika posisi centroid baru tidak sama.

Euclidian Distance

Euclidean distance adalah perhitungan jarak dari 2 buah titik dalam Euclidean space. Euclidean space diperkenalkan oleh Euclid, seorang matematikawan dari Yunani sekitar tahun 300. untuk mempelajari hubungan antara sudut dan jarak. Euclidean ini berkaitan dengan Teorema Phytagoras dan biasanya diterapkan pada 1, 2 dan 3 dimensi. Tapi juga sederhana jika diterapkan pada dimensi yang lebih tinggi.

Pada 1 dimensi. Semisal ingin menghitung jarak Euclidean 1 dimensi. Titip pertama adalah 4, titik kedua adalah -10. Caranya adalah kurangkan -10 dengan 4. sehingga menghasilkan -14. Cari nilai absolut dari nilai -14 dengan cara memangkatkannya sehingga mendapat nilai 196. Kemudian diakarkan sehingga mendapatkan nilai 14. Sehingga jarak euclidean dari 2 titik tersebut adalah 14.

Sedangkan pada 2 dimensi caranya hampir sama. Misalkan titik pertama mempunyai kordinat (1,2). Titik kedua ada di kordinat (5,5). Caranya adalah kurangkan setiap kordinat titik kedua dengan titik yang pertama. Yaitu, (5-1,5-2) sehingga menjadi (4,3). Kemudian pangkatkan masing-masing sehingga memperoleh (16,9).

Kemudian tambahkan semuanya sehingga memperoleh nilai $16+9 = 25$. Hasil ini kemudian diakarkan menjadi 5. Sehingga jarak euclidean-nya adalah 5. Euclidian distance adalah metrika yang paling sering digunakan untuk menghitung kesamaan dua vektor. Jarak euclidian dapat dianggap sebagai jarak yang paling pendek antar dua poin-poin dan pada dasarnya sama halnya dengan persamaan Phytagoras. Persamaan Euclidian distance yaitu (Urip, Adi, & Widodo, 2017):

$$\sqrt{\dots\dots\dots}(2.3)$$

Pada persamaan 2.3, dimana, adalah jarak antara data ke-i dengan centroid ke-k, adalah centroid awal untuk variable 1, adalah pusat cluster 1 untuk variable 1, adalah centroid awal untuk variable 2, adalah pusat cluster 1 untuk variable 2.

3. Hasil Penelitian Dan Pengujian

Pembersihan data

Berikut pada tabel 4.1 merupakan data asli curah hujan di wilayah JABODETABEK yang diambil dari website BMKG.

Pengelompokan K-Means

Adapun tahapan pengelompokan algoritma *K-Means* pada data cuaca JABODETABEK yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Menentukan nilai k. k=2
2. Mengambil 5 *instance* untuk dijadikan sebagai *centroid* awal yang dipilih secara acak dari tabel 4.3.
3. Dari setiap *instance* data pada tabel 4.3 diukur jarak menggunakan *Euclidean Distance* terhadap setiap *instance centroid* awal.
4. Dari perhitungan jarak *centroid* pada *cluster 1* dan *cluster 2* awal akan digunakan untuk mengelompokkan setiap *instance* data yang memiliki jarak *centroid* terdekat, sehingga diperoleh *cluster* sementara.
5. Dari *cluster* yang diperoleh akan digunakan untuk menentukan *centroid* baru dengan cara mengambil nilai rata-rata setiap indikator pada tabel 4.3 dengan *cluster* yang sama.
6. Dari setiap *instance* data pada tabel 4.3 diukur jarak menggunakan *Euclidean Distance* terhadap setiap *instance centroid* baru.
7. Dari perhitungan jarak *centroid* pada *cluster 1* dan *cluster 2* baru akan digunakan untuk mengelompokkan setiap *instance* data yang memiliki jarak *centroid* terdekat, sehingga diperoleh *cluster* akhir.
8. Hasil *cluster* yang diperoleh dari tahapan nomor 4 dan nomor 7 akan dibandingkan. Apabila *cluster* yang dihasilkan sama, maka proses pengelompokan selesai. Akan tetapi apabila *cluster* yang dihasilkan berbeda, maka proses akan dilanjutkan dengan mengulang dari tahapan nomor 2.

K-Means Dengan Euclidean Distance

Dengan menggunakan data pada tabel 4.3, akan dikelompokkan kedalam 2 kelompok menggunakan algoritma *K-Means* dengan metode perhitungan jarak *Euclidean Distance*

Tabel 4.5. Hasil dari Iterasi 1

Iterasi pertama			
data ke -i	C1	C2	Cluster
1	43.2391	40.33237	2
2	0	81.12447	1
3	90.47055	87.15538	2
4	13.25707	80.21428	1
5	80.30592	15.05457	2
6	81.09365	0	2
7	90.34644	13.63598	2

8	80.6142	16.51636	2
9	81.20179	5.175906	2
10	81.02043	5.262129	2

Dari tabel 4.5 akan digunakan untuk menentukan *centroid* baru dengan cara mengambil nilai rata-rata setiap indikator pada tabel 4.3 dengan *cluster* yang sama. Sehingga diperoleh *centroid* baru seperti yang ditunjukkan pada tabel 4.6.

cluster 1											
data ke-i	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	
1	0	0	0	0	0	0	0	0	0	0	0
2	25.2	31.6	28.2	81	7.7	0	2	5	360	2	
3	0	0	0	0	0	0	0	0	0	0	0
4	26.6	32.5	29.4	72	0	5.5	2	4	360	2	
5	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0
jml data	51.8	64.1	57.6	153	7.7	5.5	4	9	720	4	
jml data ada	2	2	2	2	1	1	2	2	2	2	
Rata-Rata	25.9	32.05	28.8	76.5	7.7	5.5	2	4.5	360	2	

Tabel 4.6. Centroid baru pada Iterasi 1 cluster 1

Tabel 4.7. Centroid baru pada Iterasi 1 cluster 2

cluster 2											
data ke-i	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	
1	26.6	33.8	30	69	0	7.3	3	6	320	1	
2	0	0	0	0	0	0	0	0	0	0	0
3	22.8	32	28.2	84	82.9	3.6	2	5	310	1	
4	0	0	0	0	0	0	0	0	0	0	0
5	24.7	31.4	27.7	85	5	3.9	2	5	280	1	
6	25.6	31.7	28.8	71	2.3	5.7	5	8	280	1	
7	25.4	30.6	28	78	0.9	0.4	4	6	270	1	
8	25.6	30.2	26.4	86	0	0	4	7	280	1	
9	25.4	32.6	29	70	0	2.6	4	5	280	1	
10	26	32.3	29	73	0	3.5	3	5	280	1	
jml data	202.1	254.6	227.1	616	91.1	27	27	47	2300	8	
jml data ada	8	8	8	8	4	7	8	8	8	8	
Rata-Rata	25.2625	31.825	28.3875	77	22.775	3.857142857	3.375	5.875	287.5	1	

Dari setiap *instance* data pada tabel 4.3 diukur jarak terhadap setiap *instance centroid* baru pada tabel 4.6 menggunakan *Euclidean distance*. Perhitungan jarak *centroid* awal pada cluster 1 baru sama seperti saat awal menentukan *centroid* Iterasi 1. Apabila *cluster* yang dihasilkan dari *centroid* awal dan *centroid* baru sama, maka perhitungan selesai.

Hasil Pengujian

Pengujian metode perhitungan jarak *Euclidean Distance* pada algoritma *K- Means* dalam mengelompokan curah hujan di daerah JABODETABEK akan divalidasi menggunakan *Davies Bouldin Index* dapat dilihat pada tabel 4.8.

Tabel 4.8. Hasil validasi DBI

Atribut	Cluster - 0	Cluster -1
i1	26.1	25.6072917
i2	32.3631579	31.628125
i3	28.9754386	28.2125
i4	74	78.0520833
i5	5.71754386	10.0604167
i6	5.20175439	4.00208333
i7	2	2.40625
i8	4.66666667	5.28125
i9	51.5789474	287.916667
i10	7.19298246	4.80208333
Jumlah	237.796491	477.96875
Rata-Rata	23.7796491	47.796875

Hasil pengujian *Davies Douldin Index* dapat dilihat pada tabel 4.8. dengan keterangan sebagai berikut :

1. Menghasilkan 2 *cluster* yaitu *cluster* 0 dan *cluster* 1
2. Jumlah *centroid* pada *cluster* 0 yaitu 237.796491 dan *cluster* 1 berjumlah 477.96875.
3. Rata-rata *centroid* pada setiap *cluster* yaitu, *cluster* 0 memiliki rata-rata 23.7796491 dan *cluster* 1 memiliki rata-rata 47.796875.

Dengan menggunakan DBI suatu *cluster* akan dianggap memiliki skema *clustering* yang optimal jika memiliki *Index Davies* minimal. Dari pengujian diatas maka nilai *Index Davies* yang dihasilkan adalah Dari pengujian metode pengelompokan data cuaca JABODETABEK tahun 2017 menggunakan algoritma *K-Means* dengan metode perhitungan jarak *Euclidean Distance* yang telah dilakukan maka dapat dilihat seperti.:

1. Berupa nilai *centroid* untuk setiap *cluster* dari metode metode perhitungan jarak *Euclidean Distance*.
2. Terbagi menjadi 2 *cluster*, dimana *cluster* 0 menunjukkan curah hujan rendah sedangkan *cluster* 1 menunjukkan curah hujan tinggi.
3. Dari perhitungan di atas maka dihasilkan rentang waktu terjadinya curah hujan tinggi yaitu pada bulan Desember sampai dengan Maret dan curah hujan rendah pada bulan April sampai dengan November, sedangkan dibulan Desember curah hujan kembali tinggi.

4. Kesimpulan dan Saran

Kesimpulan

Berdasarkan pembahasan pada bab-bab sebelumnya mengenai pengelompokan data curah hujan di JABODETABEK menggunakan algoritma *K-Means* maka dapat ditarik kesimpulan sebagai berikut:

1. Data curah hujan dapat dikelompokan menggunakan algoritma *K- Means* dan perhitungan jarak *Euclidean Distance* menjadi 2 kelompok yaitu, curah hujan tinggi dan rendah.
2. Menurut perhitungan yang sudah dilakukan pada data curah hujan di JABODETABEK maka pada bulan Januari sampai dengan Maret curah hujan tinggi dan curah hujan rendah pada bulan April sampai dengan November, sedangkan dibulan Desember curah hujan kembali tinggi.

Saran

Berdasarkan kesimpulan yang disebutkan tentang pengelompokan curah hujan maka penulis menyampaikan

saran sebagai berikut: untuk penelitian selanjutnya diharapkan dapat digunakan untuk mendeteksi kondisi alam lainnya.

Daftar Pustaka

1. Abdillah, G., Putra, F. A., & Renaldi, F. (2016). PENERAPAN DATA MINING PEMAKAIAN AIR PELANGGAN UNTUK MENENTUKAN KLASIFIKASI POTENSI PEMAKAIAN AIR PELANGGAN BARU DI PDAM TIRTA RAHARJA MENGGUNAKAN ALGORITMA K-MEANS.
2. Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016), 498-506.
3. Akbari, I. H., Novianty, A., & Setianingsih, C. (2017). Analisis Sentimen Menggunakan Metode Learning Vector Quantization. *e-Proceeding of Engineering* : Vol.4, No.2, 2283-2291.
4. Anggara, M., Sujiani, H., & Nasution, H. (2016). Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokan Member Di Alvaro Fitness. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)* Vol. 1, No. 1, 2.
5. Athoillah, i., Mariana, R., & D, E. (2017). ANALISIS SPASIAL EL NINO KUAT TAHUN 2015 DAN LA NINA LEMAH TAHUN 2016 (Pengaruhnya
6. Terhadap Kelembapan, Angin dan Curah Hujan di Indonesia). *Jurnal Sains & Teknologi Modifikasi Cuaca* vol 18, 1.
7. Chandra, A. (2016). PEMBANGKITAN ITEMSET UNTUK ATURAN ASOSIASI DENGAN ALGORITMA APRIORI DATA MINING. *Teknologi Data*
8. *Mining*, 1.
9. Dhanachandra, N., Mangleam, K., & Chanu, Y. J. (2015). Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science* 54, 764-770.
10. Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic Review on Educational Data Mining. *Digital Object Identifier* vol.5, 3.
11. Huda, N. M. (2010). APLIKASI DATA MINING UNTUK MENAMPILKAN INFORMASI TINGKAT KELULUSAN MAHASISWA. Semarang:
12. <http://eprints.undip.ac.id>.
13. Maesaroh, S., & Kusriani. (2017). Sistem Prediksi Produktifitas Pertanian Padi Menggunakan Data Mining. *Jurnal Teknologi Informasi* Vol. 7 No. 2, 26.
14. Mentari, M., Ginardi, H., & Faticah, C. (2015). SEGMENTASI PENYAKIT PADA CITRA DAUN TEBU MENGGUNAKAN FUZZY C MEANS – SUPPORT
15. VECTOR MACHINE DENGAN FITUR WARNA. *Jurnal Ilmiah Teknologi Informasi* Vol.13, No.1, 46.
16. Mutia, A. C., Sundoro, A. F., & Yajidin, A. (2017). REVIEW PENERAPAN FUZZY LOGIC SUGENO DAN MAMDANI PADA SISTEM PENDUKUNG KEPUTUSAN PRAKIRAAN CUACA DI INDONESIA. *Seminar Nasional Sistem Informasi Indonesia*, 116.
17. Nasari, F., & Darma, S. (2015). PENERAPAN K-MEANS CLUSTERING PADA
18. DATA PENERIMAAN MAHASISWA BARU. *Seminar Nasional Teknologi Informasi dan Multimedia*, 73-
19. Nasution, M. (2017). IMPLEMENTASI DATA MINING K-MEANS UNTUK MENGUKUR KEMAMPUAN LOGIKA MAHASISWA. *Jurnal Ilmiah*
20. *AMIK Labuhan Batu* Vol.5 No.1, 33.
21. Nasution, R. Y., & Eka, M. (2018). PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA APLIKASI MENENTUKAN BERAT BADAN
22. *IDEAL. Jurnal Ilmu Komputer dan Informatika* Vol.2, No.1, 78.
23. Novianti, T., & Aziz, A. (2015). APLIKASI DATA MINING MENGGUNAKAN METODE DECISION TREE UNTUK MENAMPILKAN LAPORAN HASIL NILAI AKHIR MAHASISWA (STUDI KASUS DI FAKULTAS TEKNIK UMSURABAYA) . *Jurnal Ilmiah NERO* Vol. 1 No. 3, 198-203.
24. Nugroho, R. P., & Fatta, a. H. (2012). TEKNOLOGI AUGMENTED REALITY SEBAGAI MEDIA INFORMASI PEMASARAAN PADA KUPU-KUPU MALAM CAR AUTO-FASHION. *JURNAL DASI* Vol. 13 No. 4, 29.
25. Puspitasari, N., & Hafiluddin. (2016). Penerapan Metode K-Means Dalam Pengelompokan Curah Hujan Di Kalimantan Timur. *Seminar Nasional Riset Ilmu Komputer (SNRIK 2016)* Vol.1, 20.
26. Puspitasari, N., & Havilludin. (2016). PENERAPAN METODE K-MEANS DALAM PENGELOMPOKKAN CURAH HUJAN DI KALIMANTAN TIMUR.