**TUGAS MAKALAH**

**MATAKULIAH SIMULASI PEMODELAN**

**ALGORITMA CLUSTERING**



Di Susun Oleh :

10116260 Ary Sugiarto

**UNIVERSITAS KOMPUTER INDONESIA**

**TEKNIK INFORMATIKA**

**2019**

**KATA PENGANTAR**

Assalamu’alaikum warahmatullahi wabarakatuh

Segala puji bagi Allah SWT yang telah memberikan kami kemudahan sehingga kami dapat menyelesaikan makalah ini dengan tepat waktu. Tanpa pertolongan-Nya tentunya kami tidak akan sanggup untuk menyelesaikan makalah ini dengan baik. Shalawat serta salam semoga terlimpah curahkan kepada baginda tercinta kita yaitu Nabi Muhammad SAW yang kita nanti-natikan syafa’atnya di akhirat nanti.

Penulis mengucapkan syukur kepada Allah SWT atas limpahan nikmat sehat-Nya, baik itu berupa sehar fisik maupun akal pikiran, sehingga penulis mampu untuk menyelesaikan pembuatan makalah sebagai tugas akhir dari mata kuliah Simulasi Pemodelan dengan judul “Algoritma Clustering”.

Penulis tentu menyadari bahwa makalah ini masih jauh dari kata sempurna dan masih banyak terdapat kesalahan serta kekurangan di dalamnya. Untuk itu, penulis mengharapkan kritik serta saran dari pembaca untuk makalah ini, supaya makalah ini nantinya dapat menjadi makalah yang lebih baik lagi. Demikian, dan apabila terdapat banyak kesalahan pada makalah ini penulis mohon maaf yang sebesar-besarnya.

Penulis juga mengucapkan terima kasih kepada semua pihak khususnya kepada dosen matakuliah Simulasi Pemodelan Universitas komputer Indonesia yang telah membimbing kami dalam menulis makalah ini.

Demikian, semoga makalah ini dapat bermanfaat. Terima kasih.

Bandung, 18 Maret 2019

# DAFTAR ISI

[BAB 1 PENDAHULUAN 4](#_Toc3814112)

[1.1 Latar Belakang 4](#_Toc3814113)

[1.2 Rumus Masalah 4](#_Toc3814114)

[1.3 Maksud dan tujuan 4](#_Toc3814115)

[BAB II PEMBAHASAN 5](#_Toc3814116)

[1.4 Pengertian Clustering 5](#_Toc3814117)

[1.5 Contoh penerapan algoritma Clustering K-Means 7](#_Toc3814118)

[1.6 Kegunaan Clustering 9](#_Toc3814119)

[Kesimpulan 10](#_Toc3814120)

[Daftar Pustaka 11](#_Toc3814121)

# BAB 1 PENDAHULUAN

## Latar Belakang

Teknologi informasi saat ini sudah berkembang begitu cepat dan penggunaan nya sudah menyeluruh ke berbagai sektor industri dari pendidikan, kesehatan, perusahaan dan pemerintahan sudah menerapkan kecanggihan teknologi. di era 4.0 ini teknologi yang lagi populer adalah Mesin Learning atau yang biasa disebut AI (Artificial Intelegent), teknologi yang sangat canggih dimana mesin yang di program tidak harus selalu membutuhkan inputan dari manusia untuk melakukan aksi nya, sesuai namanya mesin learning yaitu dimana mesin yg di program akan belajar sendiri untuk melakukan aksinya. hal tersebut tidak luput dari yang namanya Algoritma yan kegunaan nya bermacam-macam seperti filter,pengelompokan data dan pada makalah ini penulis akan menjelaskan algoritma yang biasa dipakai di untuk mengelompokan data yaitu algoritma Clustering.

## Rumus Masalah

1. Apa itu Algoritma Clustering?
2. Bagaimana penerapannya dan cara kerja algoritma?
3. Mengidentifikasi kegunaan Algoritma Clustering?

## Maksud dan tujuan

1. Menjelaskan apa itu algoritma clustering.
2. Menjelaskan cara kerja algoritma.
3. mengetahui kegunaan algoritma clustering.

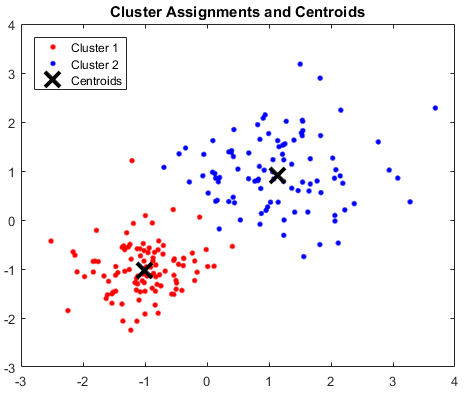
# BAB II PEMBAHASAN

## Pengertian Clustering

Clustering atau klasterisasi adalah metode pengelompokan data. Menurut Tan, 2006 clustering adalah sebuah proses untuk mengelompokan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum.

Algoritma pengelompokan (*clustering*) semacam K-Means dikenal sebagai algoritma pembelajaran tidak terbimbing (*unsupervised learning*) karena tidak adanya target kelas untuk setiap data. Algoritma K-Means menjadi salah satu algoritma yang peling penting dalam bidang data mining. Memiliki kelebihan sebagai algoritma yang mudah diimplementasikan, relatif cepat ditinjau dari waktu komputasi dan telah digunakan secara luas untuk menyelesaikan berbagai persoalan komputasi.

Penting untuk diketahui bahwa, K yang dimaksudkan dalam K-Means adalah jumlah partisi, sehingga algoritma ini mengelompokkan setiap titik pada data X dalam salah satu partisi K. Semakin besar nilai K berdampak pada banyaknya cluster untuk mengelompokkan data X. Nilai K ini juga menjadi parameter yang dibutuhkan oleh algoritma K-Means. Tidak ada ketetapan mutlak bagaimana menentukan nilai K yang optimal. Biasanya, penentuan nilai K didasarkan atas informasi yang diketahui sebelumnya tentang seberapa banyak cluster data yang muncul pada data X. Cara lain yang paling naif dalam menentukan nilai K adalah melakukan percobaan dengan beberapa nilai K hingga ditemukan bentuk pengelompokan yang tepat. Namun dewasa ini, banyak peneliti telah mencoba untuk melakukan pencarian nilai K yang optimum pada K-Means menggunakan berbagai pendekatan meta-heuristik seperti particle swarm optimization (pso), ant colony optimization (aco) dan lain sebagainya.



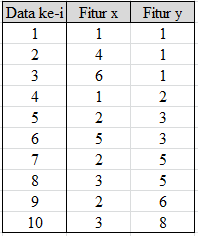
Istilah yang lazim ditemukan dalam algoritma K-Means adalah centroid. Centroid dapat diartikan sebagai titik pusat cluster, banyaknya centroid juga bergantung dari banyaknya nilai K yang diberikan. Gambar 1 menunjukkan terdapat data set yang dilakukan pengelompokan menjadi dua cluster. Maka, untuk mendapatkan dua cluster, nilai K yang diberikan adalah 2. Kemudian, dua cluster tersebut harus memiliki centroid sebagai titik pusat cluster. Penentuan centroid ini dapat dilakukan secara random atau mengambil beberapa titik data sebagai centroid. Data akan ter kelompok berdasarkan centroid terdekat, sehingga terbentuklah suatu pengelompokan data sesuai jumlah K yang diberikan.

Secara garis besar, algoritma K-Means Clustering dijelaskan dalam 5 tahap berikut :

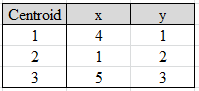
1. Inisialisasi, tentukan nilai K sebagai jumlah cluster. Jika perlu tetapkan ambang batas perubahan fungsi objektif (batas yang menentukan iterasi berhenti atau tidak) dan ambang batas perubahan posisi centroid.
2. Pilih K data dari data set X sebagai centroid.
3. Alokasikan semua data ke centroid terdekat dengan menghitung metrik jarak.
4. Hitung kembali centroid C berdasarkan data yang mengikuti cluster masing-masing.
5. Ulangi langkah 3 dan 4 hingga kondisi konvergen tercapai, yaitu (a) perubahan fungsi objektif sudah di bawah ambang batas atau (b) tidak ada data yang berpindah cluster atau (c) perubahan posisi centroid sudah berada di bawah ambang batas.

## Contoh penerapan algoritma Clustering K-Means

Pada contoh ini, disediakan data sampel berupa 10 data sebagaimana yang terpampang pada gambar 2. Data tersebut memiliki dua dimensi (fitur x dan fitur y) agar mudah di visualisasikan dalam koordinat kartesius.

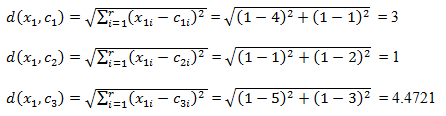


Berdasarkan data set tersebut, dilakukan proses pengelompokan menjadi 3 cluster (k = 3). berdasarkan k=3, maka ditentukan titik centroid sebanyak k berdasarkan titik-titik tertentu data set. Dapat dilakukan secara random ataupun langsung ditentukan. Maka dipilih data ke-2, 4 dan 6 sebagai centroid awal. Pengukuran jarak pada setiap data terhadap titik centroid dilakukan menggunakan perhitungan jarak Euclidean. Setelah nilai K dan centroid terinisialisasi, maka perlu ditentukan pula nilai fungsi objektif sebagai ambang batas iterasi yaitu sebesar 0.1. Sebagai permulaan, misalnya ditentukan nilai fungsi objektif sebesar 1000. Jadi jika pada iterasi tertentu nilai fungsi objektif didapatkan sudah berada di bawah ambang batas 0.1, maka iterasi akan berhenti.

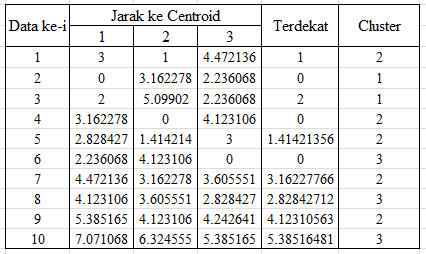


Setelah nilai K dan centroid terinisialisasi, maka perlu ditentukan pula nilai fungsi objektif sebagai ambang batas iterasi yaitu sebesar 0.1. Sebagai permulaan, misalnya ditentukan nilai fungsi objektif sebesar 1000. Jadi jika pada iterasi tertentu nilai fungsi objektif didapatkan sudah berada di bawah ambang batas 0.1, maka iterasi akan berhenti.

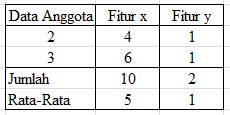
Proses perhitungan dimulai dari iterasi 1. Awalnya, dihitung jarak dari setiap data (1-10) terhadap semua centroid yang ada. Dari hasil perhitungan jarak antara setiap data terhadap semua centroid, didapatkan nilai jarak terkecil terhadap satu centroid, maka centroid tersebut disebut sebagai centroid terdekat, dan data akan berafiliasi menjadi cluster dari centroid terdekat. Berikut ini contoh menghitung jarak antara data ke-1 dengan centroid 1,2 dan 3 menggunakan metode euclidean.



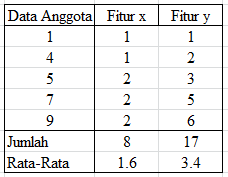
Dari perhitungan diatas, dapat diketahui bahwa data ke-1 lebih dekat jaraknya terhadap centroid ke-2. Oleh karena itu data ke-1 mengikuti cluster ke-2. Berdasarkan contoh diatas, ditentukan jarak terdekat pada semua data terhadap centroid. Jarak terdekat menentukan suatu data akan masuk ke cluster 1,2 atau 3.



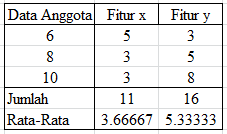
Setelah dipastikan bahwa data 1 hingga data 10 sudah masuk ke salah satu cluster. Selanjutnya, ditentukan centroid baru berdasarkan data yang tergabung pada setiap clusternya. Dimulai dari cluster 1, terdapat dua data yang tergabung didalamnya (lihat gambar 2). Centroid baru didapatkan dari rata-rata yang diperoleh yaitu 5 dan 1.



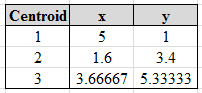
Kemudian untuk cluster ke-2, ada 5 data yang tergabung didalamnya (lihat gambar 2). Centroid baru didapatkan dari rata-rata yang diperoleh yaitu 1.6 dan 3.4.



Yang terakhir adalah untuk cluster ke-3, ada 3 data yang tergabung didalamnya (lihat gambar 2). Centroid baru didapatkan dari rata-rata yang diperoleh yaitu 3.66667 dan 5.33333.



Rata-rata yang didapatkan dari ketiga cluster tersebut merupakan centroid baru yang didapat. Centroid baru digunakan untuk perhitungan pada iterasi ke-2.



## Kegunaan Clustering

Clustering sangat berguna dan bisa menemukan group atau kelompok yang tidak dikenal dalam data. Clustering banyak digunakan dalam berbagai aplikasi seperti misalnya pada business inteligence, pengenalan pola citra, web search, bidang ilmu biologi, dan untuk keamanan (security). Di dalam business inteligence, clustering bisa mengatur banyak customer ke dalam banyaknya kelompok. Contohnya mengelompokan customer ke dalam beberapa cluster dengan kesamaan karakteristik yang kuat. Clustering juga dikenal sebagai data segmentasi karena clustering mempartisi banyak data set ke dalam banyak group berdasarkan kesamaannya.

# Kesimpulan

Kesimpulannya Clustering merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Objek yang di dalam cluster memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan cluster yang lain. Partisi tidak dilakukan secara manual melainkan dengan suatu algoritma clustering.

# DAFTAR PUSTAKA

|  |  |
| --- | --- |
| [1] | S. EDY IRWANSYAH, "BINUS," [Online]. Available: http://socs.binus.ac.id. |
| [2] | A. Fauzan, "Informatika Kita," 2017. [Online]. Available: http://www.charisfauzan.net/2016/12/algoritma-k-means-clustering-dan.html. |
| [3] | A. Fauzan, "Informatika kita," 2017. [Online]. Available: http://www.charisfauzan.net/2016/12/algoritma-k-means-clustering-dan-contoh.html. |