**MAKALAH ALGORITMA K-Mean**

*Disusun untuk memenuhi salah satu tugas mata kuliah Rekayasa Perangkat Lunak II*

*dengan dosen Alif Finandhita, S.Kom., M.T.*

**Oleh :**

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
| **Nama/NIM** | **:** | **Diar Gumilang Suhendi** | **10114253** |
|  |  | **Muhamad Solahudin** | **10114279** |
|  |  | **Muhammad Faris Somantri** | **10114252** |
|  |  | **Wildhan Ery** | **10113** |
| **Kelas** | **:** | **AI-7** | |

****

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK DAN ILMU KOMPUTER**

**UNIVERSITAS KOMPUTER INDONESIA**

**2017**

# **KATA PENGANTAR**

Puji dan Syukur kita panjatkan ke hadirat Tuhan yang Maha Esa, karena atas berkat dan hidayah-Nya saya bisa menyelesaikan tugas ini dengan baik. Tugas ini dikerjakan untuk memenuhi salah satu tugas mata kuliah Artificial Intellegent dengan judul “Makalah Algoritma K-Mean”.

Didalam Tugas ini terdapat banyak kekuranagan, untuk itu saya ingin mengucapkan permohonan maaf yang sebesar-besarnya. Kritik dan saran akan saya terima sebagai suatu masukan yang baik untuk saya di kedepannya. Tidak lupa saya sampaikan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu terselesaikannya tugas ini, mudah-mudahan semua bantuannya di berikan balasan yang terbaik oleh Tuhan Yang Maha Esa.

Untuk itu, sekali lagi saya ucapkan maaf yang sebesar-besarnya, mudah-mudahan tugas ini bermanfaat bagi kita semua. Terima kasih.

Bandung, 15 Oktober 2017

***Penyusun***

# **DAFTAR ISI**

[KATA PENGANTAR i](#_Toc495831171)

[DAFTAR ISI ii](#_Toc495831172)

[DEFINISI 1](#_Toc495831173)

[1.1 Pengertian 1](#_Toc495831174)

[1.2 Karakteristik 2](#_Toc495831175)

[1.3 Jenis-jenis 3](#_Toc495831176)

[CONTOH KASUS 5](#_Toc495831177)

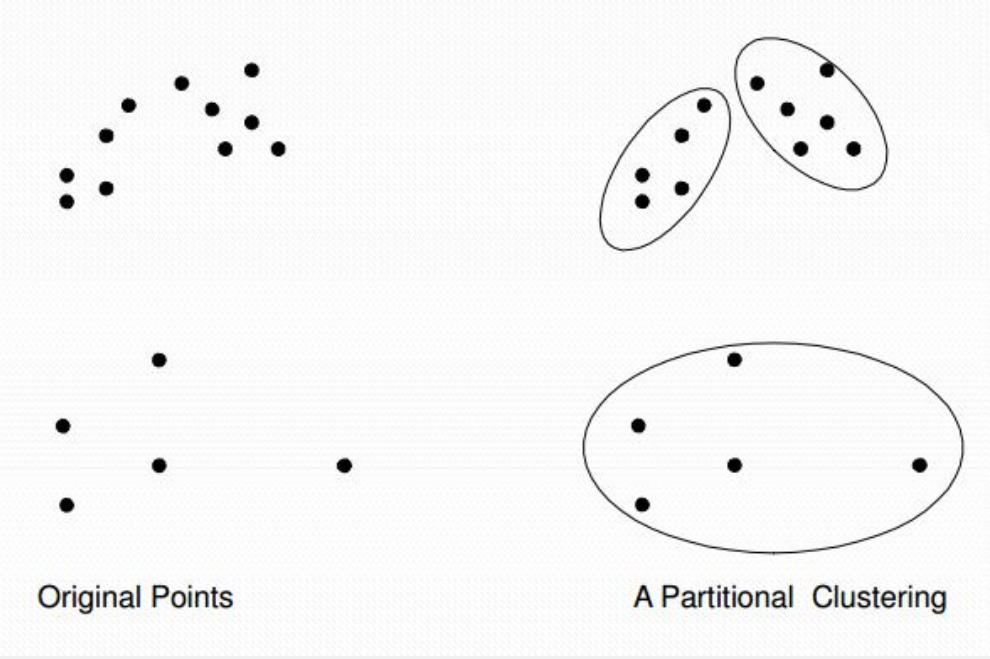
[DAFTAR PUSTAKA](#_Toc495831178)

# **DEFINISI**

## **Pengertian**

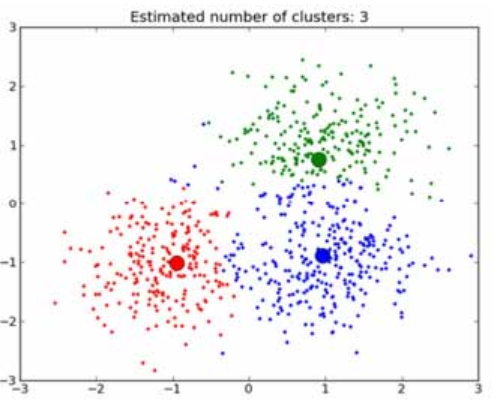
K-Means Clustering adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi.

Terdapat dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu **hierarchical** dan **non-hierarchical**, dan **K-Means** merupakan salah satu metode data clustering non-hierarchical atau **partitional clustering.**



Gambar 1 Partitional Clustering

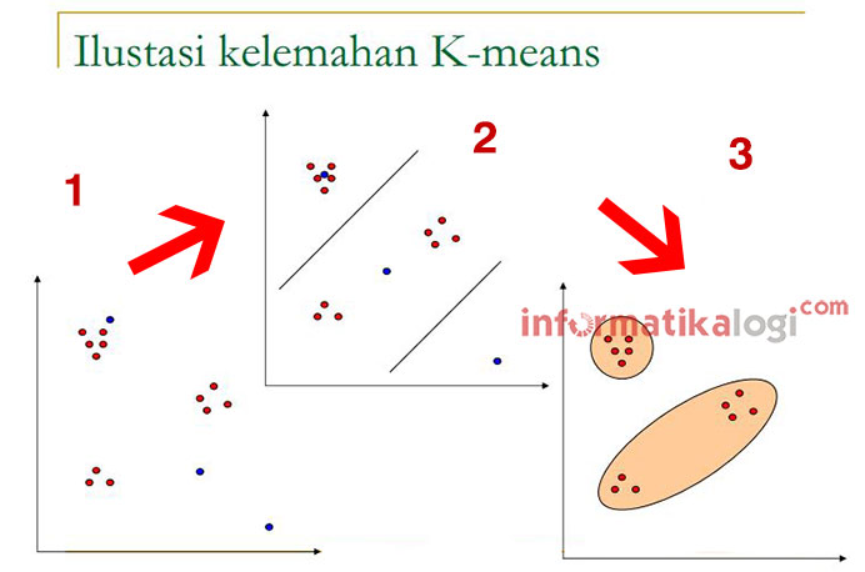
Dengan kata lain, metode K-Means Clustering bertujuan untuk meminimalisasikan objective function yang diset dalam proses clustering dengan cara meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.



Gambar 2 Contoh Cluster

## **Karakteristik**

1. K-Means sangat cepat dalam proses clustering
2. K-Means sangat sensitif pada pembangkitan centroid awal secara random
3. Memungkinkan suatu cluster tidak mempunyai anggota
4. Hasil clustering dengan K-Means bersifat tidak unik (selalu berubah-ubah) – terkadang baik, terkadang jelek
5. K-means sangat sulit untuk mencapai global optimum



Gambar 3 Ilustrasi Kelemahan K-Means

## **Jenis-jenis**

Secara mendasar, ada dua cara pengalokasian data kembali ke dalam masing-masing cluster pada saat proses iterasi clustering. Kedua cara tersebut adalah pengalokasian dengan cara tegas (hard), dimana data item secara tegas dinyatakan sebagai anggota cluster yang satu dan tidak menjadi anggota cluster lainnya, dan dengan cara fuzzy, dimana masing-masing data item diberikan nilai kemungkinan untuk bisa bergabung ke setiap cluster yang ada. Kedua cara pengalokasian tersebut diakomodasikan pada dua metode Hard K-Means dan Fuzzy K-Means. Perbedaan di antara kedua metode ini terletak pada asumsi yang dipakai sebagai dasar pengalokasian.

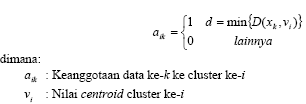
Hard K-Means

Pengalokasian kembali data ke dalam masing-masing cluster dalam metode Hard K-Means

didasarkan pada perbandingan jarak antara data dengan centroid setiap cluster yang ada.

Data dialokasikan ulang secara tegas ke cluster yang mempunyai centroid terdekat dengan

data tersebut. Pengalokasian ini dapat dirumuskan sebagai berikut[6]:

[](http://2.bp.blogspot.com/-eif-KLDa_70/TcDnN2m5MLI/AAAAAAAAADM/6By7hnG1WNM/s1600/9.PNG)

Fuzzy K-Means

Metode Fuzzy K-Means (atau lebih sering disebut sebagai Fuzzy C-Means) mengalokasikan

kembali data ke dalam masing-masing cluster dengan memanfaatkan teori Fuzzy. Teori ini

mengeneralisasikan metode pengalokasian yang bersifat tegas (hard) seperti yang digunakan

pada metode Hard K-Means. Dalam metode Fuzzy K-Means dipergunakan variabel

membership function, ik u , yang merujuk pada seberapa besar kemungkinan suatu data bisa

menjadi anggota ke dalam suatu cluster. Pada Fuzzy K-Means yang diusulkan oleh Bezdek[3],

diperkenalkan juga suatu variabel m yang merupakan weighting exponent dari membership

function. Variabel ini dapat mengubah besaran pengaruh dari membership function, ik u ,

dalam proses clustering menggunakan metode Fuzzy K-Means. m mempunyai wilayah nilai

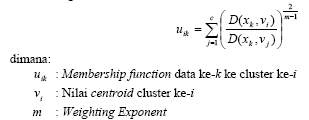
m>1. Sampai sekarang ini tidak ada ketentuan yang jelas berapa besar nilai m yang optimal

dalam melakukan proses optimasi suatu permasalahan clustering. Nilai m yang umumnya

digunakan adalah 2.

Membership function untuk suatu data ke suatu cluster tertentu dihitung menggunakan

rumus sebagai berikut[3,8,9]:

[](http://3.bp.blogspot.com/-qb-HUVm479M/TcDoY47MyFI/AAAAAAAAADQ/UmeNWHs_8Sc/s1600/3.PNG)

Membership function, ik u , mempunyai wilayah nilai 0≤ ik u ≤1. Data item yang mempunyai

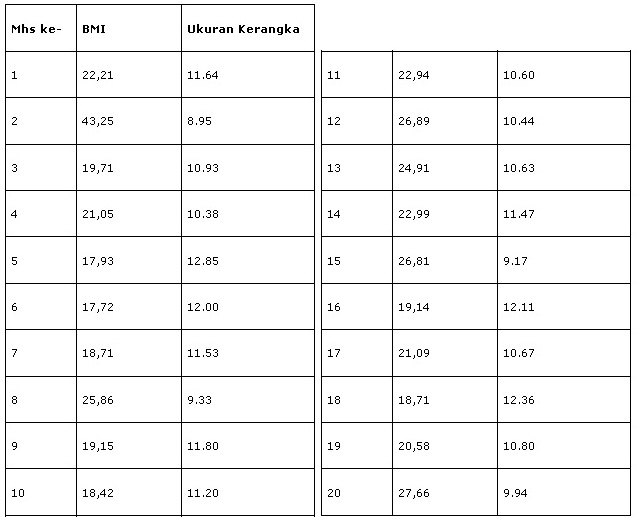
tingkat kemungkinan yang lebih tinggi ke suatu kelompok akan mempunyai nilai

membership function ke kelompok tersebut yang mendekati angka 1 dan ke kelompok yang

lain mendekati angka 0.

# **CONTOH KASUS**

Diberikan data Body Mass Index (BMI) dan ukuran kerangka 20 orang mahasiswa sebagai berikut :



Selanjutnya  kita mencoba mengelompokkan data tersebut di atas menjadi 3 kelompok. Dengan menggunakan algoritma K-Means, berikut langkah - langkah penyelesaiannya :

* Menentukan jumlah cluster, dimana jumlah cluster = 3
* Mentukan pusat cluster secara acak. Disini kita tentukan kita tentukan c1 = (20,9); c2 = (23,10); dan c3 = (27,11).
* Menghitung jarak setiap data yang ada terhadap setiap pusat cluster.  Berikut perhitungannya dengan menggunakan persamaan Euclidean Distance Space :

             - Jarak antara data mahasiswa pertama dengan pusat cluster pertama

http://smulnat.files.wordpress.com/2012/04/1.png

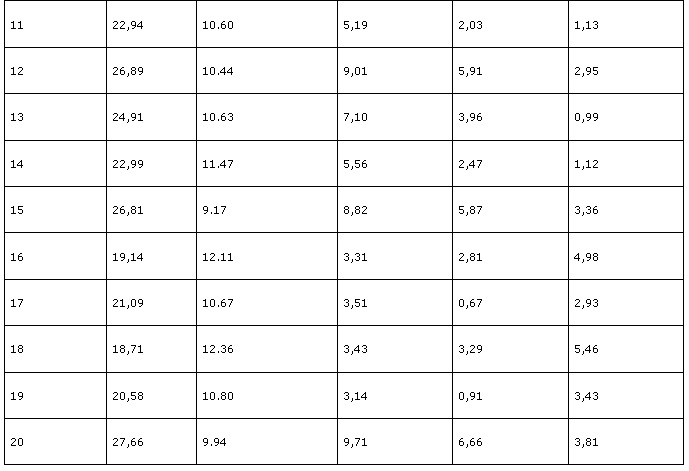
             - Jarak antara data mahasiswa pertama dengan pusat cluster ke-dua

http://smulnat.files.wordpress.com/2012/04/2.png

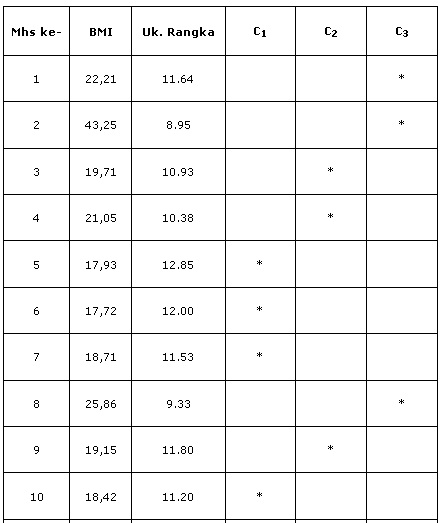
             - Jarak antara data mahasiswa pertama dengan pusat cluster ke-tiga

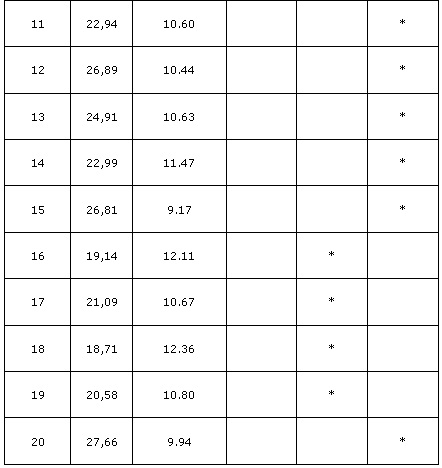
http://smulnat.files.wordpress.com/2012/04/3.png

               Adapun hasil dari perhitungan dari keseluruhan data terhadap tiap pusat cluster awal disajikan pada tabel berikut



Menentukan cluster dengan jarak terdekat pada masing-masing data. Adapun hasilnya ditampilkan pada tabel di bawah ini :





Tabel  iterasi 1

Hitung pusat cluster baru.

Untuk cluster pertama, ada 4 data yaitu data ke-5, 6, 7 dan data ke-10, sehingga:  
  
- C11 = (17,93+17,72+18,71+18,42) / 4 = **18,19**  
- C12 = (12,85+12,00+11,53+11,20) / 4 = **11,89**  
Untuk cluster kedua, ada 7 data yaitu data ke-3, 4, 9, 16, 17, 18 dan data ke-19, sehingga :  
  
- C21  = (19,71+21,05+19,15+19,14+21,09+18,71+20,58) / 7 = **19,92**  
- C22 = (10,93+10,38+11,8+12,11+10,67+12,36+10,8) / 7 = **11,29**  
  
Untuk cluster ketiga, ada 9 data yaitu data ke-1, 2, 8, 11, 12, 13, 14, 15 dan data ke-20, sehingga  
- C31 = (22,21+43,25+25,86+22,94+26,89+24,91+22,99+26,81+27,66) / 9 = **27,06**  
- C32 = (11,64+8,95+9,33+10,6+10,44+10,63+11,47+9,17+9,94) / 9 = **10,24**

* Ulangi langkah 2 (Hitung jarak setiap data yang ada terhadap setiap pusat cluster), 3 (Tentukan cluster dengan jarak terdekat pada masing-masing data) dan 4 (Hitung pusat cluster baru)hingga posisi data terhadap cluster sudah tidak mengalami perubahan.

# **DAFTAR PUSTAKA**

[1] <https://informatikalogi.com/algoritma-k-means-clustering/>. Diakses pada tanggal 15 Oktober 2017, 11.00.

[2] <http://sharewy.blogspot.co.id/2013/04/algoritma-k-means-clustering.html>. Diakses pada tanggal 15 Oktober 2017, 11.40.

[3] <http://sariberbagiilmu.blogspot.co.id/2011/05/metode-k-means.html>. Diakses pada tanggal 15 Oktober 2017, 11.50.