**K-MEANS CLUSTERING**

**PADA KELOMPOK BERAT BADAN IDEAL, OVERWEIGHT, dan UNDERWEIGHT**

**MAKALAH**

**Ditulis Untuk Memenuhi Tugas Terstruktur**

**Data Mining**



**Disusun Oleh :**

|  |  |
| --- | --- |
| Rr Octanty M | 105060800111035 |
| Anastasia Chandra U  M. Zatiar Erwan Kalam  M. Hilman Fatah  Sriningsih | 105060801111073  105060807111062  105060807111093  105060813111002 |
|  |  |
|  |  |
|  |  |

**PROGRAM TEKNOLOGI INFORMASI DAN ILMU KOMPUTER**

**UNIVERSITAS BRAWIJAYA**

**MALANG**

**2013**

**BAB I**

**PENDAHULUAN**

**1.1 Latar Belakang Masalah**

Berat badan adalah sebuah masalah dunia modern. Jika dulu, berat badan adalah cermin kemakmuran, sekarang berat badan adalah cerminan pola hidup tidak sehat/teratur, resiko tinggi menderita berbagai benyakit, serta berpengaruh pada penampilan dan gaya seseorang. Seseorang yang mempunyai berat badan yang berlebih, merasa dirinya tidak seksi, tidak sehat dan tidak menarik secara fisik. Karena zaman yang menuntut seseorang yang terlihat langsing akan percaya diri dengan penampilannya. Bahkan banyak wanita berusaha membuat tubuhnya selangsing mungkin, padahal dia sudah mengalami anoreksi nervosa, sebuah gangguan pola makan, karena keinginannya yang keras menjaga berat badannya. Bahkan seorang yang mengalami *anoreksia nervosa*, jika dilihat dari segi medis, sudah sangat membahayakan dirinya, seperti berat badannya yang tidak normal, kurus, dan tidak ada konsumsi karbohidrat, karena takut berat badannya bertambah [1].

Bila Anda ketahui di dalam ilmu pengetahuan pada dunia kedokteran bahwa antara berat badan ideal, sebenarnya tergantung dengan tinggi badan anda. Jadi berat badan ideal berbeda-beda sesuai dengan tinggi badan seseorang. Bila kita tahu cara mengukur berat badan ideal, kita jadi tahu apakan berat badan kita ideal atau tidak. Seperti kita ketahui masalah yang sering kita hadapi adalah masalah meningkatnya berat badan. Berat badan sebenarnya meningkat secara perlahan. Bila kita tidak sadar bahwa berat badan kita tidak ideal dan malah naik, kita harus waspada. Masalahnya adalah kebanyakan dari kita tidak sadar bahwa trend berat badan kita adalah naik. Yang sering terjadi kita baru sadar setelah kita mengalami kegemukan [2].

Di sini kita mencoba menguji hasil perhitungan berat badan seseorang untuk menemukan keidealan berat tubuh seseorang dengan menggunakan K-means. Pada perhitungan tersebut kita bisa melihat apakah dirinya masuk dalam kategori orang dengan berat badan ideal, kelebihan berat badan atau sudah termasuk kategori obesitas. Dari hasil tersebut juga, kamu bisa membuat rencana atau tindakan lebih lanjut untuk menyikapinya. Misalnya, bila sudah masuk dalam kategori obesitas, maka dengan cepat kamu bisa membuat program diet yang sesuai dengan kondisi tubuh. Program ini dapat digunakan rutin bagi mereka yang ingin mengetahui keidealan tubuh mereka secara rutin sehingga kesehatan dan keindahan dapat termonitor dengan baik oleh sang pemilik tubuh. Bila Proses tubuh dalam pertumbuhan dan perkembangan terpelihara dengan baik akan menunjukkan baiknya kesehatan seseorang. Seseorang yang sehat tentunya memiliki daya pikir dan kegiatan fisik sehari-hari yang cukup tinggi.

**1.2 Rumusan Masalah**

Banyaknya permasalahan yang dialami oleh seseorang dalam mengetahui keidealan tubuhnya sehingga dapat mendukung pola hidup sehat seseorang, maka diambil rumusan masalah sebagai berikut:

1. Seberapa efektifkah perhitungan untuk mengetahui keidealan tubuh seseorang dengan menggunakan perhitungan K-Means?
2. Merancangan sebuah program yang dapat membantu seseorang dalam melakukan pola hidup sehat dengan mengetahui berat badan yang ideal pada dirinya.
3. Memastikan bahwa perhitungan yang dilakukan pada program ini cukup handal untuk digunakan bagi mereka yang ingin mengetahui berat badan idealnya.
4. Pada perhitungan ini diharapkan dapat mendukung pola hidup sehat seseorang dengan mengetahui berat badan idealnya.

**1.3 Batasan Masalah**

Terdapat beberapa batasan masalah dalam membuat program untuk mengetahui berat badan ideal seseorang, yang diharapkan dapat mencapai sasaran dan tujuan. Batasan permasalahanan yang ada dibatasi sebagai berikut:

1. Program ini dibuat hanya untuk mengetahui berat badan yang ideal seseorang
2. Program ini hanya mengacu pada faktor tinggi dan berat badan seseorang untuk mengetahui keidealan tubuh seseorang.
3. Program ini melakukan perhitungan denga K-Means yang nantinya dapat terbagi sesuai dengan cluster yang sudah dipastikan oleh penulis.
4. Program dibangun menggunakan bahasa pemrograman Java.

**1.4 Tujuan**

Program ini dibuat untukmengetahui cara pengukuran berat badan ideal dengan mengukur berat badan dan tinggi badan serta nilai presentasinya. Program ini secara praktis dapat digunakan oleh semua kalangan. Tentu dengan adanya program ini akan menguntungkan bagi seseorang yang ingin mengetahui Berat Badan Ideal (BBI) sehingga orang tersebut juga secara tidak sengaja melakukan pola hidup sehat didalamnya. Karena seseorang yang melakukan pola hidup sehat ini dapat melakukan perencanaan dengan apa yang harus dilakukan untuk menunjang kesehatan tubuhnya. Dengan mengetahui berat badan ideal juga maka seseorang juga mengurangi angka obesitas yang dapat mendukung mengurangnya penyakit-pennyakit yang disebabkan oleh obesitas.

**1.5 Manfaat**

Terdapat beberapa manfaat yang diharapkan dapat memberi dampak positif baik bagi mereka yang menggunakan program ini untuk mengetahui berat badan idealnya, antara lain:

1. Bagi Pengguna:

* Hasil dari pembangunan program ini diharapkan dapat mempermudah pengguna dalam melakukan pola sehat melalui program untuk mengetahui berat badan ideal.
* Meningkatkan kinerja penghitungan dalam penghitungan untuk mengetahui berart badan ideal seseorang.
* Memajukan ilmu yang telah terkandung di dalam program ini sehingga pengguna dapat menjalankan pola hidup sehat d dalamnya

1. Bagi Pembuat program Berat Badan Ideal:

* Hasil dari penelitian berikut diharapkan dapat menambah pengetahuan Pembangun program berat badan seseorang khususnya mengenai pentingnya membangun pola hidup sehat melalui monitoring berat badan seseorang.

**BAB II**

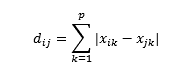
**METODE PENELITIAN**

**2.1 *K-Means Custering***

Metode *K-means Clustering* merupakan salah satu algoritma partitional clustering clustering yang paling sederhana dan umum. Hal ini dikarenakan K-*means* mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien [2]. Pada metode ini, setiap *cluster* akan dihubungkan dengan sebuah pusat cluster (centroid). Namun, K-means mempunyai mempunyai kelemahan yang diakibatkan oleh penentuan pusat awal *cluster.* Hasil *cluster* yang terbentuk dari metode K-means ini sangatlah tergantung pada inisiasi nilai pusat awal *cluster* yang diberikan. Hal ini menyebabkan hasil *cluster*nya berupa solusi yang sifatnya *local optimal*. Untuk itu, maka K-means dikolaborasikan oleh metode hierarki untuk penentuan pusat awal *cluster*. Pada permasalahan yang diangkat diartikel ini digunakan K-Means Clustering dengan 2 fitur dan 2 kelas. Dari proses pengelompokan ini nantinya diharapkan akan diketahui kemiripan atau kedekatan antar data sehingga dapat dikelompokkan ke dalam beberapa *cluster,* dimana antar anggota *cluster* memiliki tingkat kemiripan yang tinggi.

**2.2 Jarak *Manhattan (City Block/Minkowski)***

Penghitungan jarak dengan menggunakan metode *Manhattan* merupakan suatu bentuk umum dari jarak *Euclidean*, yang secara umum digunakan apabila peubah yang diamati berkolerasi atau tidak saling bebas. Dalam penggunaan metode ini, pengukuran jarak dilakukan dengan menghitung jumlah absolut perbedaan untuk tiap-tiap peubah. Secara umum, rumus penghitungan jarak menggunakan Manhattan adalah sebagai berikut:



dimana :

dij : jarak antara objek i dengan objek k

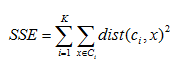
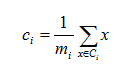
xij : nilai objek i pada peubah ke-k

xjk: nilai objek j pada peubah ke-k

P : banyaknya peubah yang diamati

**2.3 SSE (*Sum Squared Error*)**

SSE merupakan kependekan dari *Sum Squared Error*. SSE pada umumnya digunakan untuk menentukan hasil pengklasteran yang lebih baik apabila diketahui inisialisasi *centroid* atau pusat *cluster* yang berbeda-beda. SSE dirumuskan sebagai berikut:

Dari perhitungan yang telah dilakukan, maka semakin kecil nilai SE maka semakin baik. Salah satu teknik untuk memperkecil nilai SSE adalah dengan memperbesar nilai K.

**BAB III**

**PEMBAHASAN MASALAH**

**3.1 Pengertian *Clustering***

Pengertian *clustering* dalam data mining adalah suatu pengelompokan sejumlah data atau objek ke dalam *cluster* (kelompok) sehingga setiap dalam *cluster* tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek dalam *cluster* yang lainnya.

Secara umum, terdapat dua metode *clustering* yang pang banyak digunakan, yaitu *hierarchical clustering dan partitioning.* Metode *hierarchical clustering* sendiri terdiri dari *complete linkage clustering, single linkage clustering, average linkage clustering* dan *centroid linkage clustering*. Sedangkan metode *partitioning* *clustering* sendiri terdiri dari k-means dan fuzzy k-means[3].

**3.2 *Clustering***

Ada beberapa langkah-langkah pengerjaan dalam proses *clustering,*antar lain:

1. Melakukan kombinasi data sesuai dengan tabel berat badan dan tinggi badan. Tentukan jumlah klaster (*k*) yang ingin dibentuk secara acak.
2. Tentukan titik pusat *(centroid*) dari setiap *cluster*.
3. Hitung jarak objek dengan *centroid* dari tiap *cluster* yang ada menggunakan perhitungan Manhattan.
4. Objek akan dikelompokkan terhadap *cluster* dengan jarak (*d*) terpendek. Kemudian lakukan perhitungan SSE untuk mengetahui nilai error.
5. Hitung centroid yang baru dengan rumus perhitungan rata-rata.

Lanjutkan perhitungan pada iterasi berikutnya sampai nilai titik pusat *cluster* stabil (tidak berubah) [4].

**3.3 Tahap-tahap *K-Means Clustering***

1. Tentukan k sebagai jumlah *cluster* yang di bentuk.

Untuk menentukan banyaknya *cluster* k dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak *cluster*.

2. Bangkitkan k *centroid* (titik pusat *cluster*) awal secara *random*.

Penentuan *centroid* awal dilakukan secara *random*/acak dari objek-objek yang tersedia sebanyak k *cluste*r, kemudian untuk menghitung *centroid cluster* ke-i berikutnya, digunakan rumus sebagai berikut :



dimana;

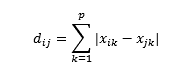
v : *centroid* pada *cluster*

x*i* : objek ke-i

n: banyaknya objek/jumlah objek yang menjadi anggota *cluster*

3. Hitung jarak setiap objek ke masing-masing centroid dari masing-masing *cluster*.

Untuk menghitung jarak antara objek dengan centroid penulis menggunakan penghitungan jarak *Manhattan.*



dimana :

dij : jarak antara objek i dengan objek k

xij : nilai objek i pada peubah ke-k

xjk: nilai objek j pada peubah ke-k

P : banyaknya peubah yang diamati

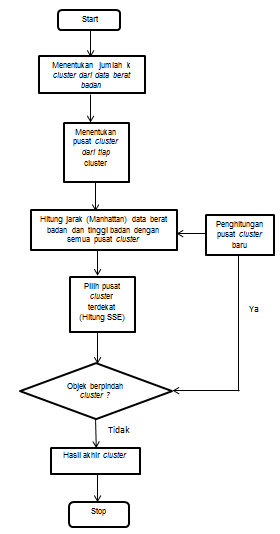
4. Alokasikan masing-masing objek ke dalam *centroid* yang paling terdekat.

5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan seperti diawal.

6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama [5].

**3.4 Flowchart *K-Means Clustering***

Berikut ini merupakan tahapan pengerjaan suatu proses *clustering* dengan menggunakan *K-Means Clustering* yang digambarkan dalam bentuk flowchart:



* 1. Pengerjaan

Berikut ini merupakan urutan pengerjaan :

* Langkah awal dengan mencari beberapa berat badan dan tinggi badan setiap orang, dan parameter ideal, overweight dan underweight berdasarkan BMI (Body Mass Index) :

BMI < 18,5         = Under weight (berat badan kurang)

BMI 18,5 – 24,9  = Normal

BMI 25 – 29,9     = Overweight (berat badan berlebih )

BMI > 30            = Obese (gemuk )

Berikut ini adalah sample beberapa data dari berat badan setiap orang

|  |  |  |  |
| --- | --- | --- | --- |
| NO. | NAMA | BERAT | TINGGI |
| (kg) | (cm) |
| 1 | Adhe Fitri | 87 | 154 |
|
| 2 | Andi Hariati | 50 | 154 |
|
| 3 | Anwar Amir | 50 | 163 |
|
| 4 | Asmar | 39 | 167 |
|
| 5 | Eka Andriyani | 45 | 156 |
|
| 6 | Evi Multazam | 31 | 145 |
|

1. Kita akan menggunakan dua fitur (tinggi badan dan berat badan) untuk mengetahui apakah termasuk dalam kategori ideal, overweight dan underweight seperti data yang terdapat pada table dibawah ini :

|  |  |  |  |
| --- | --- | --- | --- |
| NO. | NAMA | BERAT | TINGGI |
| (kg) | (cm) |
| 1 | Adhe Fitri | 87 | 154 |
|
| 2 | Andi Hariati | 50 | 154 |
|
| 3 | Anwar Amir | 50 | 163 |
|
| 4 | Asmar | 39 | 167 |
|
| 5 | Eka Andriyani | 45 | 156 |
|
| 6 | Evi Multazam | 31 | 145 |
|

1. Data yang ada akan dikelompokkan menjadi 3 kelompok (ideal, overweight, underweight). Adapun langkah dari *pengelompokkan* data adalah sebagai berikut :

*Data pada iterasi pertama.* Tentukan pusat cluster secara acak, misal kita tentukan m1 = (37,148); m2 = (52,155); dan m3 = (80,157 )*.* Hitung jarak setiap data yang ada terhadap setiap pusat cluster. Untuk menghitung jarak data tinggi badan, berat badan dengan pusat cluster pertama menggunakan persamaan (1) (dengan perhitungan manhattan) adalah :

* Menentukan jarak Cluster pertama dengan berat dan tinggi badan adhe fitri :

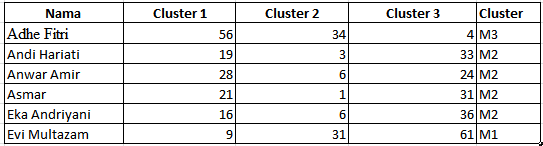
= |(87-37)+(154-148)| = 56

* Menentukan jarak Cluster kedua dengan berat dan tinggi badan adhe fitri:

= |(87-52)+(154-155)| = 34

* Menentukan jarak Cluster ketiga dengan berat dan tinggi badan adhe fitri

= |(87-80)+(154-157)| = 4



Suatu data akan menjadi anggota dari suatu cluster yang memiliki jarak terkecil dari pusat clusternya. Misalkan untuk data pertama, jarak terkecil diperoleh pada cluster pertama, sehingga data pertama akan menjadi anggota dari cluster pertama. Demikian juga untuk data kedua, jarak terkecil ada pada cluster kedua, maka data tersebut akan masuk pada cluster kedua.

*Data pada iterasi kedua*

1. Hitung pusat cluster baru untuk cluster pertama, ada data 1 yaitu data ke 6 :

M1(berat badan):

31/1=31

M1(tingi badan):

145/1 = 145

Jadi M1 = (31,145 )

1. Hitung pusat cluster baru untuk cluster kedua, ada data 4 yaitu data ke 2,3,4,5 :

M2(berat badan):

(50 + 50 + 39 + 45)/4 = 46

M2(tingi badan):

(154+163+167+156)/4=160

Jadi M2 = (46,160 )

1. Hitung pusat cluster baru untuk cluster pertama, ada data 1 yaitu data ke 1 :

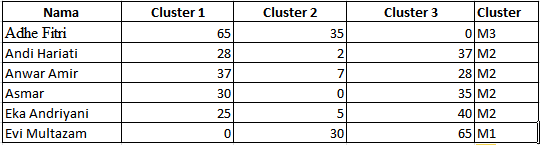
M3(berat badan):

87/1= 87

M3(tingi badan):

154/1=154

Jadi M3 = (87,154 )



Karena pada iterasi kedua posisi centroid tidak berubah sehingga proses pengklasteran telah selesai.

BAB IV

PENJELASAN PROGRAM

Penjelasan program:

Pada program awal mula diminta inputan berat badan dan tinggi badan sebanyak data 6 orang. Kemudian terjadi penghitungan manhattan dengan masing-masing data tersebut



Math.abs= berfungsi untuk menghitung nilai mutlak

xp1, yp1 sebagai titik pusat cluster pertama

xp2, yp2 sebagai titik pusat cluster kedua

xp3,yp3 sebagai titik pusat cluster ketiga

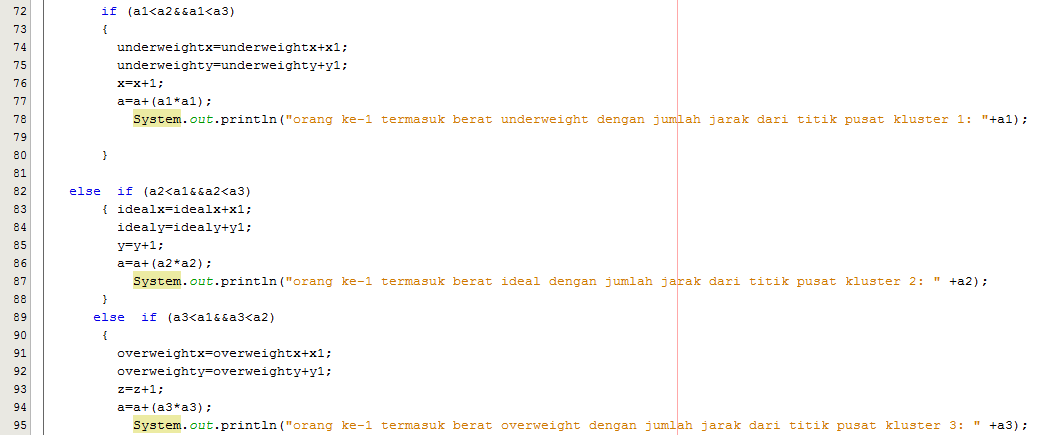
x1 sebagai fitur berat badan badan , dengan urutan x1 sampai x6

y2 sebagai fitur tinggi badan , dengan urutan y1 sampai x1

a1: untuk melakukan perhitungan x dan y dengan xp1 dan yp1

a2: untuk melakukan perhitungan x dan y dengan xp2 dan yp2

a3: untuk melakukan perhitungan x dan y dengan xp3 dan yp3



Kemudian terjadi percabangan pada program , dari perhitungan manhattan tersebut akan masuk ke cluster yang mana, program akan mengecek hasil perhitungan dengan nilai paling terkecil, apabila :

* nilai perhitungan a1 paling kecil maka akan masuk ke cluster 1
* nilai perhitungan a2 paling kecil maka akan masuk ke cluster 2
* nilai perhitungan a3 paling kecil maka akan masuk ke cluster 3

untuk mendapatkan nilai pusat kluster baru dari jumlah fitur tinggi badan dan berat badan yang masuk ke cluster masing-masing

sehingga pada masing-masing percabangan terdapat syntax



Syntax tersebut berguna untuk menambahkan jumlah variable nilai tinggi badan dan badan yang masuk ke cluster underweight



Syntax tersebut berguna untuk menambahkan jumlah variable nilai tinggi badan dan badan yang masuk ke cluster ideal



Syntax tersebut berguna Untuk menambahkan jumlah variable nilai tinggi badan dan badan yang masuk ke cluster ideal

Untuk menjumlahkan jumlah data yang masuk ke tiap cluster yaitu dengan syntax:

untuk menambahkan jumlah data yang masuk ke cluster underweight

 untuk menambahkan jumlah data yang masuk ke cluster ideal

 untuk menambahkan jumlah data yang masuk ke cluster overweight

Sehingga penghitungan titik cluter baru tiap cluster (xp dan yp) baru adalah sebagai berikut

Untuk cluster underweight



Pusat cluster baru untuk cluster underweight menjadi (m1a,m1b)

Untuk cluster ideal



Pusat cluster baru untuk cluster underweight menjadi (m2a,m2b)

Untuk cluster overweight



Pusat cluster baru untuk cluster underweight menjadi (m3a,m3b)

Perhitungan nilai sse dengan pemasukkan nilai a ke variable SSE



Dimana nilai a didapat dari hasil penjumlahan kuadrat, apakah itu kuadrat dari a1, a2, atau a3, dilihat dari nilai yang paling kecil, hasil nilai yang paling kecil akan dikuadratkan dan masing-masing akan ditambahkan kemudian masuk ke variable a.

Syntaxnya adalah sebagai berikut pada percabangan





}

Untuk nilai kuadrat a1 yang dimasukkan ke nilai a



{



}

Untuk nilai kuadrat a2 yang dimasukkan ke nilai a



{



}

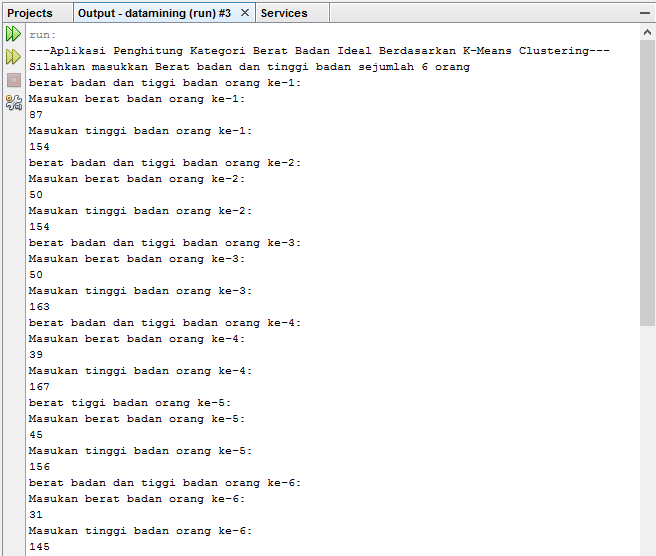
**BAB V**

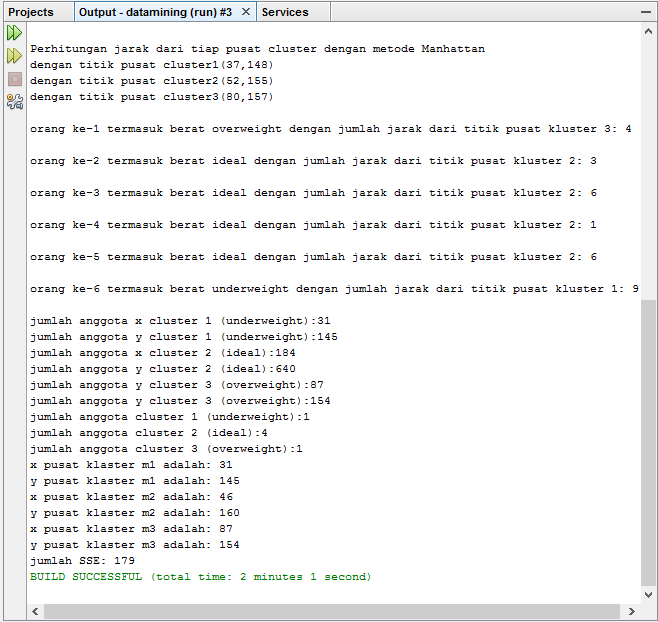
KESIMPULAN

1. Kesimpulan

Hasil masing-masing pusat klaster dari clusternya adalah Dari hasil percobaan diatas dapat disimpulkan bahwa algoritma K-Means dapat digunakan untuk mengelompokkan data dari berat badan seseorang apakah underweight, ideal atau overweight diatas berdasarkan tinggi badan dan berat badan seseorang.

Hasil Screen Shot Output Aplikasi:





1. Saran
2. Penulis mengharapkan terdapat pengembangan lebih lanjut dari aplikasi, untuk pemasukkan data set yang lebih banyak.
3. Penulis adanya pengembangan lebih lanjut terhadap aplikasi ini dengan dapat menginputkan lebih banyak fitur sehingga aplikasi dapat membuat keputusan yang lebih tepat .

DAFTAR PUSTAKA

1. Garda Pengetahuan. *Tips Menentukan Berat Badan Ideal*. 2010. URL: <http://garda-pengetahuan.blogspot.com/2012/04/tips-menentukan-berat-badan-ideal.html>,diakses pada tanggal 8 November 2013.
2. Anonymous. *Manfaat Mengetahui Berat Badan Ideal Seseorang*. 2013. URL: <http://dietaman.com/manfaat-mengetahui-berat-badan-ideal/>, diakses pada tanggal tanggal 8 November 2013.
3. Anonymous. *Rumus Berat Badan Ideal Seseorang*. 2013. URL: <http://carakata.org/berat-badan-ideal-rumus-cara-menghitung-dan-mengukur-berat-badan/>, diakses pada tanggal tanggal 8 November 2013.
4. Ediyanto, Novitasari Mara Muhlasah, Satyahadewi Neva.2013. “*Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis*”.Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster) Volume 02 , No. 2 (2013), hal 133 – 136, jurnal.untan.ac.id
5. Agusta, Yudi. Pebruari 2007. “*K-Means- Penerapan, Permasalahan dan Metode Terkait*”. Jurnal Sistem dan Informatika Vol.3 : 47-60.
6. Septiya Hindriyana Eka Nanda,Mula’ab, Khusnul Khotimah Bain**. “**Pemanfaatan Madm ( *Multiple Attribute Decission Making* ) Pada Pemilihan Mahasiswa Berprestasi Menggunakan Algoritma Dia (*The Distance To The Ideal Alternative*)”. 2013.URL: [http://pta.trunojoyo.ac.id/uploads/journals/090411100001/090411100001.pdf](http://pta.trunojoyo.ac.id/uploads/journals/090411100001/090411100001.pdf" \t "_blank) diakses pada 5 Juni 2013