Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 75-88
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Penerapan Algoritma K-Means Dalam Mengelompokkan Rata-Rata Konsumsi Kalori Menurut Provinsi

Juwitha Lovely Sweets Sinaga[1], Solikhun[2], Dedi Suhendro[3]
[1]Mahasiswa Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar
[2,3]AMIK Tunas Bangsa, Pematangsiantar
[1]juwithalovelysweetssinaga@gmail.com, [2]solikhun@amiktunasbangsa.ac.id,
[3]dedi.su@amiktunasbangsa.ac.id

Abstract

Calories are a source of energy that we get from food intake that contains nutrients and as a basic human need for humans to survive. The number of people who consume excessive calories and do not pay attention to the amount of calorie intake that is consumed will result in the emergence of various diseases that are bad for health. In this case the government does not have information about the data on the average calorie consumption per province by province. The purpose of this study is to determine the highest and lowest clusters, for that the authors use Data Mining with the K-Means Algorithm to classify Average Calorie Consumption per day by province. This test is carried out using RapidMiner software. The results were obtained from the average grouping of calorie consumption grouped by two clusters: high and low clusters, high clusters of 13 provinces and low clusters of 21 provinces. Provinces that are classified as low clusters are expected to be a contribution for the Indonesian government for decision making in an effort to maintain a balanced consumption of calories per day and create a healthy lifestyle program in the future.

Keywords*: S K-Means, Clustering, RapidMiner, Rata-rata konsumsi kalori*

Abstrak

Kalori adalah sumber energi yang kita dapatkan dari asupan makanan yang mengandung nutrisi dan sebagai kebutuhan pokok manusia agar manusia dapat bertahan hidup. Banyaknya masyarakat yang mengkonsumsi kalori secara berlebihan dan tidak memperhatikan jumlah asupan kalori yang dimakan akan mengakibatkan munculnya berbagai macam penyakit yang berakibat buruk bagi kesehatan. Dalam hal ini pemerintah belum memiliki informasi tentang data pengelompokkan rata-rata konsumsi kalori per hari menurut provinsi. Tujuan dari penelitian ini adalah untuk mengetahui cluster teringgi dan rendah, untuk itu penulis menggunakan Data Mining dengan Algoritma K-Means untuk mengelompokkan Rata-rata Konsumsi Kalori per hari menurut provinsi. Pengujian ini dilakukan dengan menggunakan software RapidMiner. Hasil penelitian diperoleh dari pengelompokkan rata-rata konsumsi kalori yang dikelompokkan dengan dua cluster yaitu cluster tinggi dan rendah, cluster rendah 13 provinsi dan cluster tinggi 21 provinsi. Provinsi yang termasuk cluster rendah diharapkan dapat dijadikan kontribusi bagi pemerintah Indonesia pemerintah untuk pengambilan keputusan dalam upaya menjaga keseimbangkan konsumsi kalori per harinya dan membuat program pola hidup sehat di masa yang akan datang.

Kata kunci*: K-Means, Clustering, RapidMiner, Rata-rata konsumsi kalor*

1. PENDAHULUAN

Kalori merupakan salah satu nutrisi yang terkandung dalam makanan. Kebutuhan energi seseorang menurut FAO/WHO (1985) adalah konsumsi energi berasal dari makanan yang diperlukan untuk menutupi pengeluaran energi seseorang [1]. Kalori merupakan salah satu kebutuhan pokok bagi tubuh manusia agar dapat bertahan hidup dan fungsi untuk memicu tubuh agar bisa bernafas dan

bergerak. Ada beberapa macam makanan pokok dan minuman yang mengandung jumlah kalori tinggi, seperti nasi yang mengandung karbohidrat, kacang-kacangan, buah alpokat, sirup, es krim dan lainnya. Makanan yang dimakan jika dibarengi dengan aktifitas sehari-hari akan membuat tubuh menjadi stabil. Rata-rata kebutuhkan kalori tergantung dari berat badan, jenis kelamin, usia dan aktivitas. Seiring bertambahnya usia maka metabolisme tubuh akan berjalan dengan sangat lambat. Sehingga membuat energi yang kita butuhkan berkurang dibandingkan saat usia masih muda. Kebutuhan kalori tubuh perharinya 1800 kkal sampai 2000 kkal. Total kebutuhan energi harian yang harus dikonsumsi tidak boleh kurang dari BMR yang ditetapkan. BMR bergantung terhadap parameter berat badan(BB), tinggi badan (TB), umur (U), dan jenis kelamin (JK) dan jenis aktivitas yang dilakukan [2].

Meningkatnya jumlah masyarakat dalam mengkonsumsi makanan yang mengandung kalori tinggi setiap harinya dapat mengakibatkan timbulnya beberapa penyakit seperti penyakit diabetes, stroke, penyakit jantung dan dapat mengalami obesitas. Tubuh yang mengalami obesitas yaitu masalah pada kegemukan, dimana berat badan terlalu berlebihan dan tidak seimbang akan sangat beresiko. Namun sebaliknya, jika tubuh kekurangan kalori tubuh akan merasa lemas tidak bergairah, karna tidak adanya energi bahkan membuat tubuh menjadi kurus dan dapat mengalami gizi buruk. Dalam hal ini pemerintah belum memiliki informasi tentang data pengelompokkan rata-rata konsumsi kalori per hari menurut provinsi. Data tersebut sangat dibutuhkan pemerintah sebagai informasi agar dapat memberikan arahan kepada masyarakat, untuk menjaga kesehatan dengan melakukan upaya menjaga keseimbangkan dalam mengkonsumsi kalori per harinya dan membuat program pola hidup sehat dengan berolahraga.

Setiap asupan makanan yang di konsumsi, masyarakat sebaiknya mengetahui bagaimana cara mengolah makanan dan tidak mengkonsumsi makanan secara berlebihan setiap harinya, sehingga makanan yang kita makan tidak mengakibatkan timbulnya berbagai macam penyakit dan obesitas. Dengan dilakukannya penelitian ini diharapkan dapat menjadi informasi bagi pemerintah agar lebih memperhatikan masyarakat dan memberikan fasilitas program pola hidup sehat dengan berolahraga. Perlunya olahraga minimal dua kali seminggu dan aktifitas yang memerlukan pergerakan banyak dapat membuat tubuh menjadi ideal dan seimbang, dan untuk masyarakat yang kekurangan kalori pemerintah dapat memberikan bantuan berupa asupan makanan bergizi dengan kalori yang seimbang. Maka solusi yang tepat menurut penulis dalam melakukan penelitian ini adalah dengan menggunakan Data *Mining* Metode K-*means* untuk mengelompokkan rata-rata konsumsi kalori. Menurut S. Agustina, dkk (dalam [3] 'K-Means merupakan suatu algoritma yang digunakan dalam pengelompokkan secara pertisi yang memisahkan data ke dalam kelompok yang berbeda – berbeda. Algoritma ini mampu meminimalkan jarak antara data ke clusternya'.

Diharapkan penelitian ini dapat memberikan kontribusi kepada pemerintah, ahli kesehatan dan masyarakat dalam upaya menentukan rata-rata konsumsi kalori.

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 75-88
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

2. METODOLOGI PENELITIAN

2.1. Data Mining

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Data mining mulai ada sejak 1990-an sebagai cara yang benar dan tepat untuk mengambil pola dan informasi yang digunakan untuk menemukan hubungan antara data untuk melakukan pengelompokkan ke dalam satu atau lebih cluster sehingga objek - objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. Data mining merupakan bagian dari proses penemuan pengetahuan dari basis data Knowledge Discovery in [4].

2.2. RapidMiner

RapidMiner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari *Institute of Technologi Blanchardstown dan Ralf Klinkenberg* dari rapid-i.com dengan tampilan GUI *(Graphical User Interface)* sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini [5].

2.3. Algoritma K-Means

K-Means adalah pengelompokan data non hirarki (sekatan) yang mempartisi data ke dalam bentuk dua atau lebih kelompok sehingga data yang berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama [6].
Langkah-langkah melakukan clustering dengan metode K-Means penjelasannya adalah sebagai berikut [7]:

a) Algoritma k-means menerima inputan dari pengguna berupa nilai jumlah klaster sebanyak k klaster, dan suatu dataset yang ingin dikelompokkan.
b) Sebanyak k data dipilih secara random dari suatu dataset yang kemudian ditentukan sebagai initial centroid.
c) Ulang langkah 4, 5 dan 6 hingga tidak ada lagi objek yang berubah di dalam suatu klaster.
d) Hitung jarak masing-masing titik objek data ke masing-masing intial centroid.
e) Mengalokasikan masing-masing titik data ke dalam klaster dengan jarak paling minimum.
f) Menghitung rata-rata dari semua data yang terdapat di dalam klaster tersebut sebagai pusat klaster baru.

3. HASIL DAN PEMBAHASAN

Data yang digunakan untuk penelitian ini yaitu data rata-rata konsumsi kalori menurut provinsi mulai dari tahun 2007- 2018 dan data ini di peroleh dari BPS (Badan Pusat Statistika). Perangkat lunak yang digunakan sebagai alat pengujinya yaitu *Rapidminer* 5.3 untuk menyajikan *cluster* terendah dan tertinggi yang paling produktif. Data yang diperoleh di Menentukkan jumlah data yang akan di *cluster*, dimana sampel data jumlah rata-rata konsumsi kalori yang kan digunakan dalam proses *clustering* adalah data jumlah yang terdapat pada Badan

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 75-88
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Pusat Statistik pada tahun 2007 sampai 2018 dengan jumlah data sebanyak 34 provinsi. Berikut adalah beberapa cara untuk mencari nilai rata-rata.

R1 = (2210,02 + 2080,83 + 2010,37 + 2075,79 + 1962,62 + 1869,93 + 1823,36 + 1794,04 + 1989,61 + 2021,08 + 2115,09 + 2101,57)/12 = 2004,53

R2 = (2069,52 + 2074,76 + 1921,70 + 1970,81 + 1993,59+ 1892,36+ 1848,80 + 1883,81 + 2010,98 + 1999,99 + 2133,84 + 2184,97)/12 = 1998,76

Sampai R34 dan berikut hasil dari nilai rata-rata untuk *cluster* pertama: transformasikan ke format data *Microsoft excel* 2010. Kumpulan data sebagai data masukkan dalam membuat model aturan algoritma *Clustering* K-*Means* menggunakan aplikasi *Rapidminer*.

3.1. Pengolahan Data Manual K-*means*

Untuk mendapatkan hasil dari penelitian yang dilakukan, berikut uraian perhitungan manual proses clustering dilakukan mulai dari penentuan data yang ingin di *cluster*. Dalam hal ini variabel data yang ingin di *cluster* adalah data dari tahun 2007 sampai 2018 dan disini peneliti mengambil nilai rata-rata dari tahun 2007 sampai 2018 sebagai data yang akan di *cluster*. Berikut adalah langkah-langkah penyelesaian yang dilakukan penulis dalam mengelompakkan rata-rata konsumsi kalori menurut provinsi menggunakan algoritma K-*means*.

Tabel 1. Rata-rata Konsumsi Kalori

No	Nama	Nilai Rata-Rata
1	Aceh	2004,53
2	Sumatera Utara	1998,76
3	Sumatera Barat	2075,46
4	Riau	2001,57
5	Jambi	1962,46
6	Sumatera Selatan	2023,63
7	Bengkulu	2021,29
8	Lampung	1977,69
9	Kepulauan Bangka Belitung	1966,93
...	...	...
34	Papua	1831,95

a) Pada tahap ini penulis menetapkan nilai k ada sebanyak 2 (dua) *cluster* yang akan diterapkan dalam perhitungan manual K-*means* yaitu *cluster* tinggi (*maximum*) dan rendah (*minimum*) supaya dapat menjadi kontribusi kepada pemerintah Indonesia untuk memperhatikan jumlah provinsi yang mengkonsumsi kalori agar tidak meningkat di tahun-tahun berikutnya.
b) Menentukan nilai *centroid* (pusat *cluster*) awal yang telah ditentukan secara *random* berdasarkan nilai *variabel* data yang di *cluster* sebanyak yang ditentukan sebelumnya. Adapun *cluster* tertinggi di peroleh dari nilai dan *cluster* rendah diambil dari nilai terkecil pada tabel 2.
Cara mencari nilai *centroid* data awal untuk *iterasi* 1 adalah dengan mencari nilai tertinggi dari nilai rata-rata seperti contoh dibawah ini:

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 75-88
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

C1 = 2004,53/1998,76/2075,46/2001,57/1962,46/2023,63/2021,29/1977,69/1966,93/2020,61/1950,29/2011,52/1921,56/1950,21/1932,06/2044,12/2182,63/2048,68/1894,85/1930,01/2060,30/2098,41/1834,52/1875,67/2032,60/1987,71/2047,42/1990,46/1936,70/1989,29/1836,66/1773,51/1816,21/1831,95 = 2182,63

C2 = 2004,53/1998,76/2075,46/2001,57/1962,46/2023,63/2021,29/1977,69/1966,93/2020,61/1950,29/2011,52/1921,56/1950,21/1932,06/2044,12/2182,63/2048,68/1894,85/1930,01/2060,30/2098,41/1834,52/1875,67/2032,60/1987,71/2047,42/1990,46/1936,70/1989,29/1836,66/1773,51/1816,21/1831,95 = 1773,51

Adapun hasil dari data *centroid* data awal seperti tabel di bawah ini.

Tabel 2. *Centroid* Data Awal (*Iterasi* 1)

Cluster	Nilai
Tinggi (*Max*/C1)	2182,63
Rendah (*Min*/C2)	1773,51

c) Menghitung jarak setiap data rata-rata konsumsi kalori terhadap pusat *cluster*. Setelah nilai pusat *cluster* awal ditentukan, maka langkah selanjutnya adalah menghitung jarak masing-masing data terhadap pusat *cluster* dengan menggunakan rumus yang perhitungannya dapat kita lihat sebagai berikut:

D(C1.1) = $\sqrt{(2182{,}63 - 2004{,}53)^2}$ = 178,11

Dan seterusnya Sampai D(C1.34) setelah itu lanjut ke *centroid cluster* ke dua seperti berikut ini.

D(C2.1) = $\sqrt{(1773{,}51 - 2004{,}53)^2}$ = 231,01

Dan seterusnya Sampai D(C2.34)

Berikut tabel *iterasi* 1 dapat dilihat pada tabel 4.3 dibawah ini:

Tabel 3. *Iterasi* 1 Rata-rata Konsumsi Kalori

No	Nama	Nilai Rata-Rata	C1	C2
1	Aceh	2004,53	178,11	231,01
2	Sumatera Utara	1998,76	183,87	225,25
3	Sumatera Barat	2075,46	107,17	301,95
4	Riau	2001,57	181,07	228,05
5	Jambi	1962,46	220,17	188,95
6	Sumatera Selatan	2023,63	159,00	250,12
7	Bengkulu	2021,29	161,34	247,78
8	Lampung	1977,69	204,95	204,17
9	Kepulauan Bangka Belitung	1966,93	215,71	193,41
...	...	...	...	...
34	Papua	1831,95	350,68	58,44

Setelah menghitung *centroid cluster* C1 dan C2 maka dapatlah hasil pengelompokkan jarak terpendek berdasarkan *centroid* pada *iterasi* 1 dengan

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 75-88
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

menggunakan rumus pada Ms.*Excel* = **MIN(C1;C2)** dapat kita lihat pada tabel 4 berikut ini:

Tabel 4. Jarak *Centroid Iterasi* 1

No	Nilai Rata-Rata	C1	C2	Jarak Terpendek	Hasil
1	2004,53	178,11	231,01	178,11	C1
2	1998,76	183,87	225,25	183,87	C1
3	2075,46	107,17	301,95	107,17	C1
4	2001,57	181,07	228,05	181,07	C1
5	1962,46	220,17	188,95	188,95	C2
6	2023,63	159,00	250,12	159,00	C1
7	2021,29	161,34	247,78	161,34	C1
8	1977,69	204,95	204,17	204,17	C1
9	1966,93	215,71	193,41	193,41	C2
...	...	...	...	...	...
34	1831,95	350,68	58,44	58,44	C2

d) Dalam menentukan posisi *cluster* dari tabel 4.4 diatas, dapat dilakukan dengan melihat pernyataan berikut. "jika jarak terpendek berada di kolom C1 maka pada tabel posisi *cluster*, kolom C1 diberi nilai 1" dan jika jarak terpendek berada di kolom C2 maka pada tabel posisi *cluster*, kolom C2 diberi nilai 1". Nilai 1 hanya untuk simbolis atau pertanda bahwa pada kolom tersebut terdapat nilai jarak terpendek mewakili kolom tersebut. Berikut adalah tebel posisi *cluster* pada *iterasi* 1:

Tabel 5. Posisi *cluster* pada Iterasi 1

No	Jarak Terpendek	Hasil	C1	C2
1	178,11	C1	1	
2	183,87	C1	1	
3	107,17	C1	1	
4	181,07	C1	1	
5	188,95	C2		1
6	159,00	C1	1	
7	161,34	C1	1	
8	204,17	C1	1	
9	193,41	C2		1
...	...	...	...	...
34	58,44	C2		1

Maka diperolehlah hasil dari posisi *cluster* 1 pada C1 berjumlah 19 data dan posisi *cluster* 1 pada C2 berjumlah 15 data. Dapat dilihat provinsi mana saja yang masuk pada C1 dan C2 pada tabel 6 berikut:

Tabel 6. Hasil *Iterasi* 1

Cluster	**Provinsi**	**Hasil**
C1	1,2,3,4,6,7,8,10,12,16,17,18,21,22,25,26,27,28,30	19

C2	5,9,11,13,14,15,19,20,23,24,29,31,32,33,34	15

e) Jika nilai *centroid* hasil *iterasi* dengan nilai *centroid* sebelumnya bernilai sama atau nilai *centroid* sudah optimal serta posisi *cluster* data rata-rata konsumsi kalori tidak mengalami perubahan lagi maka proses iterasi berhenti. Namun jika nilai *centroid* tidak sama atau belum optimal serta posisi data rata-rata konsumsi kalori masih berubah maka proses *iterasi* masih akan berlanjut pada *iterasi* berikutnya sampai nilai centroidnya .

Karena pada iterasi 1 nilai *centroid* tidak bernilai sama maka akan dilanjutkan pada *iterasi* ke 2 dan seterusnya hingga centriodnya bernilai sama. Untuk menghitung *Iterasi* ke-2, dapat dilakukan menggunakan cara yang sama pada *Iterasi* ke-1 yaitu sebagai berikut:

a) Pada Iterasi ke-2 yaitu menghitung jarak antara rata-rata dengan nilai k *centroid* pusat yang telah ditentukan, akan tetapi nilai k *centroid* iterasi 1 dan 2 berbeda, nilai k *centroid* pada itersi 2 di peroleh dengan menambah rata-rata hasil iterasi 1 dengan jumlah hasil pada *iterasi* 1 sebagai berikut:
C1=(2004,53+1998,76+2075,46+2001,57+2023,63+2021,29+1977,69+2020,61+2011,52+2044,12+2182,63+2048,68+2060,30+2098,41+2032,60+1987,71+2047,42+1990,46+1989,29)/19 = 2032,46
C2=1962,46+1966,93+1950,29+1921,56+1950,21+1932,06+1894,85+1930,01+1834,52+1875,67+936,70+1836,66+1773,51+1816,21+1831,95)/15 = 1894,24

b) Kemudian menghitung jarak antara rata-rata dengan nilai k *centroid* dengan menggunakan rumus perhitungan sebagai berikut:
D(C1.1) = $\sqrt{(2032{,}46 - 2004{,}53)^2}$ = 27,93
Dan seterusnya Sampai D(C1.34) setelah itu lanjut ke centroid *cluster* ke dua seperti berikut ini.
D(C2.1) = $\sqrt{(1894{,}24 - 2004{,}53)^2}$ = 110,29

Dan seterusnya Sampai D(C2.34)
Berikut tabel *iterasi* 2 dapat dilihat pada tabel 7 dibawah ini:

Tabel 7. *Iterasi* 2 Rata-rata Konsumsi Kalori

No	Nama	Nilai Rata-Rata	C1	C2
1	Aceh	2004,53	27,93	110,29
2	Sumatera Utara	1998,76	33,70	104,52
3	Sumatera Barat	2075,46	43,00	181,22
4	Riau	2001,57	30,89	107,33
5	Jambi	1962,46	70,00	68,22
6	Sumatera Selatan	2023,63	8,82	129,39
7	Bengkulu	2021,29	11,16	127,05
8	Lampung	1977,69	54,77	83,45
9	Kepulauan Bangka Belitung	1966,93	65,53	72,69
10	Kepulauan Riau	2020,61	11,85	126,37
34	Papua	1831,95	200,50	62,28

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 75-88
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

c) Setelah menghitung *centroid cluster* C1 dan C2 maka dapatlah hasil pengelompokkan jarak terpendek berdasarkan *centroid* pada *iterasi* 2 dengan menggunakan rumus pada Ms.*Excel* = **MIN(C1;C2)** dapat kita lihat pada tabel 8 berikut ini:

Tabel 8. Jarak *Centroid* Iterasi 2

No	Nilai Rata-Rata	C1	C2	Jarak Terpendek	Hasil
1	2004,53	27,93	110,29	27,93	C1
2	1998,76	33,70	104,52	33,70	C1
3	2075,46	43,00	181,22	43,00	C1
4	2001,57	30,89	107,33	30,89	C1
5	1962,46	70,00	68,22	68,22	C2
6	2023,63	8,82	129,39	8,82	C1
7	2021,29	11,16	127,05	11,16	C1
8	1977,69	54,77	83,45	54,77	C1
9	1966,93	65,53	72,69	65,53	C1
10	2020,61	11,85	126,37	11,85	C1
34	1831,95	200,50	62,28	62,28	C2

d) Dalam menentukan posisi *cluster* dari tabel 4.8 diatas, dapat dilakukan dengan melihat pernyataan berikut. "jika jarak terpendek berada di kolom C1 maka pada tabel posisi *cluster*, kolom C1 diberi nilai 1" dan jika jarak terpendek berada di kolom C2 maka pada tabel posisi *cluster*, kolom C2 diberi nilai 1". Nilai 1 hanya untuk simbolis atau pertanda bahwa pada kolom tersebut terdapat nilai jarak terpendek mewakili kolom tersebut. Berikut adalah tebel posisi *cluster* pada *iterasi* 2:

Tabel 9. Posisi *cluster* pada Iterasi 2

No	Jarak Terpendek	Hasil	C1	C2
1	27,93	C1	1	
2	33,70	C1	1	
3	43,00	C1	1	
4	30,89	C1	1	
5	68,22	C2		1
6	8,82	C1	1	
7	11,16	C1	1	
8	54,77	C1	1	
9	65,53	C1	1	
10	11,85	C1	1	
34	62,28	C2		1

Maka diperolehlah hasil dari posisi *cluster* pada C1 berjumlah 20 data dan posisi *cluster* 1 pada C2 berjumlah 14 data. Dapat dilihat provinsi mana saja yang masuk pada C1 dan C2 pada tabel 4.10 berikut:

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 75-88
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Tabel 10. Hasil Iterasi 2

Cluster	Provinsi	Hasil
C1	1,2,3,4,6,7,8,9,10,12,16,17,18,21,22,25,26,27,28,30	20
C2	5,11,13,14,15,19,20,23,24,29,31,32,33,34	14

e) Jika nilai *centroid* hasil iterasi dengan nilai *centroid* sebelumnya bernilai sama atau nilai *centroid* sudah optimal serta posisi *cluster* data rata-rata konsumsi kalori tidak mengalami perubahan lagi maka proses *iterasi* berhenti. Namun jika nilai *centroid* tidak sama atau belum optimal serta posisi data rata-rata konsumsi kalori masih berubah maka proses *iterasi* masih akan berlanjut pada *iterasi* berikutnya sampai nilai centroidnya .

Karena pada *iterasi* 1 dan 2 nilai *centroid* tidak bernilai sama maka akan dilanjutkan pada *iterasi* ke 3 dan seterusnya hingga centriodnya bernilai sama. Untuk menghitung Iterasi ke-3, dapat dilakukan menggunakan cara yang sama pada *Iterasi* ke 1 dan 2 yaitu sebagai berikut:

a) Pada *iterasi* ke-3 yaitu menghitung jarak antara rata-rata dengan nilai k *centroid* pusat yang telah ditentukan, nilai k *centroid* iterasi 1 dan 2 berbeda, tetapi nilai k *centroid iterasi* 2 dan 3 sama. nilai k *centroid* pada *iterasi* 2 dan 3 di peroleh dengan menambah rata-rata hasil *iterasi* 1 dengan jumlah hasil pada *iterasi* 2 sebagai berikut:
C1=(2004,53+1998,76+2075,46+2001,57+2023,63+2021,29+1977,69+1966,93+2020,61+2011,52+2044,12+2182,63+2048,68+2060,30+2098,41+2032,60+1987,71+2047,42+1990,46+1989,29)/20 = 2029,18
C2=1962,46+1950,29+1921,56+1950,21+1932,06+1894,85+1930,01+1834,52+1875,67+936,70+1836,66+1773,51+1816,21+1831,95)/14 = 1889,05

b) Kemudian menghitung jarak antara rata-rata dengan nilai k *centroid* dengan menggunakan rumus perhitungan sebagai berikut:
D(C1.1) = $\sqrt{(2029{,}18 - 2004{,}53)^2}$ = 224,65
Dan seterusnya Sampai D(C1.34) setelah itu lanjut ke centroid *cluster* ke dua seperti berikut ini.
D(C2.1) = $\sqrt{(1889{,}05 - 2004{,}53)^2}$ = 115,48
Dan seterusnya Sampai D(C2.34)

Berikut tabel *iterasi* 3 dapat dilihat pada tabel 11 dibawah ini:

Tabel 11. *Iterasi* 3 Rata-rata Konsumsi Kalori

No	Nama	Nilai Rata-Rata	C1	C2
1	Aceh	2004,53	27,93	110,29
2	Sumatera Utara	1998,76	33,70	104,52
3	Sumatera Barat	2075,46	43,00	181,22
4	Riau	2001,57	30,89	107,33
5	Jambi	1962,46	70,00	68,22
...	...	...	...	...
34	Papua	1831,95	200,50	62,28

c) Setelah menghitung *centroid cluster* C1 dan C2 maka dapatlah hasil pengelompokkan jarak terpendek berdasarkan *centroid* pada *iterasi* 3 dengan menggunakan rumus pada Ms.*Excel* = **MIN(C1;C2)** dapat kita lihat pada tabel 12 berikut ini:

Tabel 12. Jarak *Centroid* Iterasi 3

No	Nilai Rata-Rata	C1	C2	Jarak Terpendek	Hasil
1	2004,53	24,65	115,48	24,65	C1
2	1998,76	30,42	109,71	30,42	C1
3	2075,46	46,28	186,42	46,28	C1
4	2001,57	27,61	112,52	27,61	C1
5	1962,46	66,72	73,41	66,72	C1
...	...	...	...	...	...
34	1831,95	197,23	57,09	57,09	C2

d) Dalam menentukan posisi *cluster* dari tabel 12 diatas, dapat dilakukan dengan melihat pernyataan berikut. "jika jarak terpendek berada di kolom C1 maka pada tabel posisi *cluster*, kolom C1 diberi nilai 1" dan jika jarak terpendek berada di kolom C2 maka pada tabel posisi *cluster*, kolom C2 diberi nilai 1". Nilai 1 hanya untuk simbolis atau pertanda bahwa pada kolom tersebut terdapat nilai jarak terpendek mewakili kolom tersebut. Berikut adalah tebel posisi *cluster* pada *iterasi* 3:

Tabel 13. Posisi *cluster* pada *Iterasi* 3

No	Jarak Terpendek	Hasil	C1	C2
1	24,65	C1	1	
2	30,42	C1	1	
3	46,28	C1	1	
4	27,61	C1	1	
5	66,72	C1	1	
...	...	...	...	...
34	57,09	C2		1

Maka diperolehlah hasil dari posisi *cluster* pada C1 berjumlah 21 data dan posisi *cluster* 1 pada C2 berjumlah 13 data. Dapat dilihat provinsi mana saja yang masuk pada C1 dan C2 pada tabel 14 berikut:

Tabel 14. Hasil *Iterasi* 3

Cluster	**Provinsi**	**Hasil**
C1	1,2,3,4,5,6,7,8,9,10,12,16,17,18,21,22,25,26,27,28,30	21
C2	11,13,14,15,19,20,23,24,29,31,32,33,34	13

e) Jika nilai *centroid* hasil iterasi dengan nilai *centroid* sebelumnya bernilai sama atau nilai *centroid* sudah optimal serta posisi *cluster* data rata-rata

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 75-88
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

konsumsi kalori tidak mengalami perubahan lagi maka proses *iterasi* berhenti. Namun jika nilai *centroid* tidak sama atau belum optimal serta posisi data rata-rata konsumsi kalori masih berubah maka proses *iterasi* masih akan berlanjut pada *iterasi* berikutnya sampai nilai centroidnya .

Karena pada *iterasi* 1, 2 dan 3 nilai *centroid* tidak bernilai sama maka akan dilanjutkan pada *iterasi* ke 4 dan seterusnya hingga centriodnya bernilai sama. Untuk menghitung Iterasi ke-4, dapat dilakukan menggunakan cara yang sama pada *Iterasi* ke 1, 2 dan 3 yaitu sebagai berikut:

a) Pada *Iterasi* ke-4 yaitu menghitung jarak antara rata-rata dengan nilai k *centroid* pusat yang telah ditentukan, nilai k *centroid* iterasi 1 dan 2, 3 berbeda, tetapi nilai k *centroid* iterasi 2, 3 dan 4 sama. nilai k *centroid* pada *iterasi* 2, 3 dan 4 di peroleh dengan menambah rata-rata hasil *iterasi* 1 dengan jumlah hasil pada *iterasi* 3 sebagai berikut:
C1=(2004,53+1998,76+2075,46+2001,57+1962,46+2023,63+2021,29+1977,69+1966,93+2020,61+2011,52+2044,12+2182,63+2048,68+2060,30+2098,41+2032,60+1987,71+2047,42+1990,46+1989,29)/21 = 2026,00
C2=1950,29+1921,56+1950,21+1932,06+1894,85+1930,01+1834,52+1875,67+936,70+1836,66+1773,51+1816,21+1831,95)/13 = 1883,40

b) Kemudian menghitung jarak antara rata-rata dengan nilai k *centroid* dengan menggunakan rumus perhitungan sebagai berikut:
D(C1.1) = $\sqrt{(2026{,}00 - 2004{,}53)^2}$ = 21,48
Dan seterusnya Sampai D(C1.34) setelah itu lanjut ke centroid *cluster* ke dua seperti berikut ini.
D(C2.1) = $\sqrt{(1883{,}40 - 2004{,}53)^2}$ = 121,13
Dan seterusnya Sampai D(C2.34)

Berikut tabel iterasi 4 dapat dilihat pada tabel 15 dibawah ini:

Tabel 15. *Iterasi* 4 Rata-rata Konsumsi Kalori

No	Nama	Nilai Rata-Rata	C1	C2
1	Aceh	2004,53	21,48	121,13
2	Sumatera Utara	1998,76	27,24	115,36
3	Sumatera Barat	2075,46	49,46	192,06
4	Riau	2001,57	24,44	118,17
5	Jambi	1962,46	63,54	79,06
...	...	...	...	...
34	Papua	1831,95	194,05	51,44

c) Setelah menghitung *centroid cluster* C1 dan C2 maka dapatlah hasil pengelompokkan jarak terpendek berdasarkan centroid pada iterasi 4 dengan menggunakan rumus pada Ms.*Excel* = **MIN(C1;C2)** dapat kita lihat pada tabel 16 berikut ini:

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 75-88
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Tabel 16. Jarak *Centroid* Iterasi 4

No	Nilai Rata-Rata	C1	C2	Jarak Terpendek	Hasil
1	2004,53	21,48	121,13	21,48	C1
2	1998,76	27,24	115,36	27,24	C1
3	2075,46	49,46	192,06	49,46	C1
4	2001,57	24,44	118,17	24,44	C1
5	1962,46	63,54	79,06	63,54	C1
...	...	...	...	..	...
34	1831,95	194,05	51,44	51,44	C2

d) Dalam menentukan posisi *cluster* dari tabel 4.16 diatas, dapat dilakukan dengan melihat pernyataan berikut. "jika jarak terpendek berada di kolom C1 maka pada tabel posisi *cluster*, kolom C1 diberi nilai 1" dan jika jarak terpendek berada di kolom C2 maka pada tabel posisi *cluster*, kolom C2 diberi nilai 1". Nilai 1 hanya untuk simbolis atau pertanda bahwa pada kolom tersebut terdapat nilai jarak terpendek mewakili kolom tersebut. Berikut adalah tebel posisi *cluster* pada *iterasi* 4:

Tabel 17. Posisi *cluster* pada Iterasi 4

No	Jarak Terpendek	Hasil	C1	C2
1	21,48	C1	1	
2	27,24	C1	1	
3	49,46	C1	1	
4	24,44	C1	1	
5	63,54	C1	1	
9	59,08	C1	1	
...	...	...	...	...
34	51,44	C2		1

Maka diperolehlah hasil dari posisi *cluster* pada C1 berjumlah 21 data dan posisi *cluster* 1 pada C2 berjumlah 13 data. Dapat dilihat provinsi mana saja yang masuk pada C1 dan C2 pada tabel 18 berikut:

Tabel 18 . Hasil *Iterasi* 4

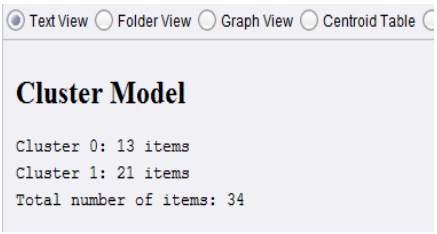
Cluster	**Provinsi**	**Hasil**
C1	1,2,3,4,5,6,7,8,9,10,12,16,17,18,21,22,25,26,27,28,30	21
C2	11,13,14,15,19,20,23,24,29,31,32,33,34	13

Proses perhitungan K-*means* berhenti pada *iterasi* ke-4, karena pada *iterasi* ke-4 sama dengan hasil *iterasi* sebelumnya yaitu *iterasi* ke-3. Diperoleh hasil dari posisi *cluster* 4 adalah C1 berjumlah 21 data dan C2 berjumlah 13 data, maka proses perhitungan tidak mengalami perubahan lagi sehingga proses *iterasi* berhenti.

Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)
Volume 6 Nomor 1 Februari, pp 75-88
ISSN: 2527-5771/EISSN: 2549-7839
https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

3.2. Hasil Percobaan

Pengujian yang dilakukan penulis menggunakan *Rapidminer* 5.3 sehingga dapat diketahui provinsi mana yang masuk kedalam *cluster* tinggi dan *cluster* rendah. Hasil pengelompokkan dua cluster pada tahap keluaran dapat dilihat dengan mengklik *Cluster Model (Clustering)* pada *toolbar*. Berikut adalah tampilan hasil *cluster* model:



Gambar 5. Nilai *cluster* Modal *Rapidminer*

4. SIMPULAN

Dari hasil pembahasan sebelumnya kesimpulan yang didapat adalah sebagai berikut:

a) Teknik penerapan data mining menggunakan algoritma k-*means* dapat diterapkan dalam meng-*cluster* jumlah provinsi rata-rata konsumsi kalori mulai tahun 2007 sampai dengan 2018.
b) Data yang sebelumnya diolah dalam perhitungan metode K-*means clustering* telah dilakukan pengujian menggunakan *software Rapidminer* versi 5.3. Sebagai pembuktian bahwa hasil dari *software Rapidminer* sama dengan hasil perhitungan manual pada metode K-*means Clustering* dan hasil yang diperoleh dalam meng-*cluster* data rata-rata konsumsi menurut provinsi yang digunakan adalah valid.
   1) Jumlah *cluster* 0 (Rendah) berjumlah 13 items yaitu provinsi DKI Jakarta, Jawa Tengah, DI Yogyakarta, Jawa Timur, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Timur, Kalimantan Utara1, Gorontalo, Maluku, Maluku Utara, Papua Barat, Papua,
   2) Jumlah *cluster* 1 (Tinggi) berjumlah 21 items yaitu provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, Jawa Barat, Banten, Bali, Nusa Tenggara Barat, Kalimantan Tengah, kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat.

DAFTAR PUSTAKA

[1] I. E. A. Novita And Hernawan Sulistyanto, "Pengembangan Aplikasi Untuk Mengetahui Kebutuhan Jumlah Kalori," 2015.
[2] M. W. Lestari Dwi Asih, "Meminimumkan Jumlah Kalori Di Dalam Tubuh Dengan Memperhitungkan Asupan Makanan Dan Aktivitas Menggunakan Linear Programming," Vol. 16, No. 1, Pp. 38–44, 2016.
[3] M. G. Sadewo, A. P. Windarto, And S. R. Andani, "Pemanfaatan Algoritma Clushtering Dalam Mengelompokkan Jumlah Desa / Kelurahan Yang Memiliki Sarana Kesehatan," Vol. I, Pp. 124–131, 2017.
[4] S. R. Ningsih, I. S. Damanik, A. P. Windarto, H. S. Tambunan, J. Jalaluddin, And A. Wanto, "Analisis K-Medoids Dalam Pengelompokkan Penduduk Buta Huruf

Menurut Provinsi," *Pros. Semin. Nas. Ris. Inf. Sci.*, Vol. 1, No. September, P. 721, 2019, Doi: 10.30645/Senaris.V1i0.78.

[5] S. Haryati, A. Sudarsono, And E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," Vol. 11, No. 2, Pp. 130–138, 2015.

[6] R. D. Kusumah *Et Al.*, "Perbandingan Metode K – Means Dan Self Organizing Map (Studi Kasus: Pengelompokan Kabupaten/Kota Di Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia 2015)," Vol. 6, Pp. 429–437, 2017.

[7] Gusti Ngurah Wisnu Paramartha, Dian Eka Ratnawati, And A. W. Widodo, "Analisis Perbandingan Metode K-Means Dengan Improved Semi-Supervised Analisis Perbandingan Metode K-Means Dengan Improved Semi- Supervised K-Means Pada Data Indeks Pembangunan Manusia (Ipm)," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. Vol. 1, No. January, Pp. 813–824, 2017.