

Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi

**Muhammad Dwi Chandra¹, Eka Irawan², Ilham Syahputra Saragih³, Agus Perdana Windarto⁴,
Dedi Suhendro⁵**

^{1,2,3,4,5}Jurusian Sistem Informasi, STIKOM Tunas Bangsa Pematangsiantar

¹muhammadchandra1998@gmail.com, ²eka.irawan@amiktunasbangsa.ac.id, ³ilham@amiktunasbangsa.ac.id,

⁴agus.perdana@amiktunasbangsa.ac.id, ⁵dedi.su@amiktunasbangsa.ac.id

Keywords:

*K-Means Algorithm,
Toddler,
BPS,
Province,
Data Mining,*

ABSTRACT

The purpose of this study was to screen toddlers who were experiencing severe malnutrition according to province. Sources of research data used were obtained from the Ministry of Health of the Republic of Indonesia. The variables used are toddlers who experience malnutrition according to the Province. In this study using Data Mining Techniques using the K-means algorithm. It is expected that the results of this study can provide input to the central government to pay more attention to nutritional intake in infants, so as to increase the growth and development of toddlers in Indonesia. . And the data obtained by high clusters are 15 Provinces yaitu (Aceh, Sumatera Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Papua Barat, Papua), and cluster rendah ada 19 yaitu (Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Dki Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Timur, Kalimantan Utara, Sulawesi utara, Maluku Utara).

Kata Kunci

*Algoritma K-Means,
Balita,
BPS,
Provinsi
Data Mining,*

ABSTRAK

Tujuan penelitian ini adalah untuk menelusuri balita yang mengalami gizi buruk menurut Provinsi. Sumber data penelitian yang digunakan diperoleh dari BPS. Adapun variabel yang digunakan adalah balita yang mengalami gizi buruk menurut Provinsi. Pada Penelitian ini menggunakan Teknik *Data Mining* dengan menggunakan algoritma *K-means*. Diharapkan hasil penelitian ini dapat memberikan masukan kepada pihak pemerintah pusat agar lebih memperhatikan asupan gizi pada balita, sehingga dapat meningkatkan pertumbuhan dan perkembangan balita yang ada di Indonesia. Dan data yang diperoleh cluster tinggi ada 15 Provinsi yaitu (Aceh, Sumatera Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Papua Barat, Papua), dan cluster rendah ada 19 yaitu (Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Dki Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Timur, Kalimantan Utara, Sulawesi utara, Maluku Utara).

Korespondensi Penulis:

Muhammad Dwi Chandra,
STIKOM Tunas Bangsa,
Jalan Sudirman Blok A No 1, 2, 3 Pematangsiantar
Telepon: +62 856 6769 4581
Email: muhammadchandra1998@gmail.com

1. PENDAHULUAN

Masa balita biasa sering dikatakan sebagai masa *golden age* atau masa keemasan karena masa pertumbuhan atau perkembangan seseorang di masa selanjutnya ditentukan pada saat balita. Pertumbuhan pada balita tidak hanya digunakan sebagai gambaran dalam bertambahnya ukuran anggota tubuh, tetapi juga digunakan sebagai gambaran

mengenai kesinambungan antara asupan dan kebutuhan gizi [1]. Gizi merupakan salah satu asupan yang sangat penting bagi pertumbuhan balita karena gizi yang baik membuat balita tumbuh normal dan jika gizi yang diperoleh balita tersebut buruk maka pertumbuhan balita dapat terhambat.

Tidak semua balita di Indonesia mendapatkan gizi yang baik. Sehingga perlu adanya pengelompokan untuk mengetahui wilayah mana yang balitanya masih kurang mendapatkan gizi baik, maka dari itu pemberian gizi harus ditingkatkan agar kebutuhan balita yang ada di Indonesia tentang gizi dapat terpenuhi dengan baik dan balita dapat berkembang dengan baik. Untuk mengetahui apakah balita tumbuh dan berkembang dengan baik adalah dengan cara melihat tumbuh kembang balita idealnya seperti apa. Kalau tinggi atau berat balita tidak ideal bisa dipastikan balita tersebut mengalami gizi buruk.

Pada penelitian ini, penulis akan membuat penelitian dengan mengelompokkan balita yang mengalami gizi buruk menurut Provinsi dengan menggunakan metode *k-means*. Adapun data yang diperoleh bersumber dari Badan Pusat Statistik (BPS). Dimana yang diperoleh adalah balita yang mengalami gizi buruk menurut Provinsi. 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 *database* yang besar [2]. Metode data *mining* memiliki beberapa bagian, salah satunya adalah *clustering*. *Clustering* memiliki berbagai metode yang digunakan untuk mengelompokkan salah satunya adalah metode *K-means*. *K-Means* merupakan salah satu metode data *clustering* non hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster/kelompok*.

Berdasarkan permasalahan yang telah diuraikan diatas, penulis berinisiatif melakukan penelitian untuk mengelompokkan balita yang mengalami gizi buruk menurut Provinsi menggunakan data *mining algoritma k-means clustering*. Dengan mengcluster balita yang mengalami gizi buruk menjadi 2 *cluster* yaitu *cluster* tinggi dan *cluster* rendah. Penelitian ini diharapkan dapat membantu Pemerintahan dalam hal gizi balita di Indonesia. Penulis memilih menggunakan metode *K-mean* ialah karena dari penelitian-panelitian terdahulu mengatakan metode *K-means* mampu bekerja dengan baik dalam mengelompokkan data dalam skala kecil maupun skala besar dengan hasil yang akurat.

Tabel 1. Penelitian Terkait

No	Judul	Penulis	Resume
1	Penerapan <i>fuzzy K-Nearest Neighbor</i> (FK-NN) dalam menentukan status gizi balita	Nugraha, D., Dalam Regasari, R., Putri, M., Wihandika, R. C [1].	Berdasarkan penelitian yang dilakukan dapat disimpulkan bahwa gizi buruk adalah keadaan kurang gizi yang disebabkan karena kekurangan asuran energi dan protein juga mikronutrien dalam jangka waktu lama.
2	Implementasi Algoritma <i>K-Means</i> untuk pemetaan Produktivitas panen padi di Kabupaten Karawang	M. Rosyid Ridlo, Sofi Defiyanti, Aji Primajaya [3].	Berdasarkan penelitian yang dilakukan dapat disimpulkan bahwa Produktivitas padi di Kabupaten Karawang dapat dipetakan menggunakan teknik pengelompokan data mining menjadi 3 kelompok, yang terdiri dari Produktivitas panen melebihi target, sesuai target dan kurang target dengan persentase kenaikan target yang ditetapkan oleh dinas pertanian Provinsi Jawa Barat sebesar 3%.
3	Analisis <i>cluster</i> dengan menggunakan metode <i>single linkage</i> dan metode <i>K-means</i>	Rezki Wahyuni, Sigit Nugroho, Pepi Novianti [4].	Berdasarkan penelitian yang dilakukan dapat disimpulkan bahwa <i>pengclusteran</i> yang dilakukan dengan menggunakan tiga data yang berbeda yaitu Produksi, Produktivitas dan luas tanaman Sayuran ini menghasilkan pemecahan <i>cluster</i> yang berbeda antara metode <i>single linkage</i> dan metode <i>K-means</i> .

2. METODE PENELITIAN

2.1 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 *database* yang besar [2].

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran *computer* (*machinelearning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis [5].

2.2 Metode K-Means

Algoritma K-Means merupakan metode *clustering* nonhierarki yang memiliki waktu komputasi yang relatif cepat. Berdasarkan analisis perbandingan antara *K-Means* dan *Fuzzy C-Means* (FCM) yang dilakukan oleh (Suomi G. dan Sanjay Kumar D), didapatkan hasil yang membuktikan bahwa algoritme K-Means lebih cepat dengan

elapsed time 0.433755 detik dibandingkan dengan algoritma FCM yang memiliki *elapsed time* sebesar 0.781679 detik [6]. Algoritma *K-Means* merupakan algoritma yang relatif sederhana untuk mengklasifikasikan atau mengelompokkan sejumlah besar obyek dengan atribut tertentu ke dalam kelompok-kelompok (*cluster*) sebanyak *K*. Pada algoritma *K-Means*, jumlah klaster *K* sudah ditentukan lebih dahulu [7].

Algoritma *K-Means* merupakan salah satu algoritma *partitional*, karena *K-Means* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya. Algoritma *K-means* menggunakan proses secara berulang untuk mendapatkan basis data *cluster*. Untuk mendapatkannya, dibutuhkan jumlah *cluster* awal yang diinginkan sebagai masukan dan penghasil jumlah *cluster* akhir sebagai *output*.

Langkah-langkah algoritma *K-Means* adalah sebagai berikut:

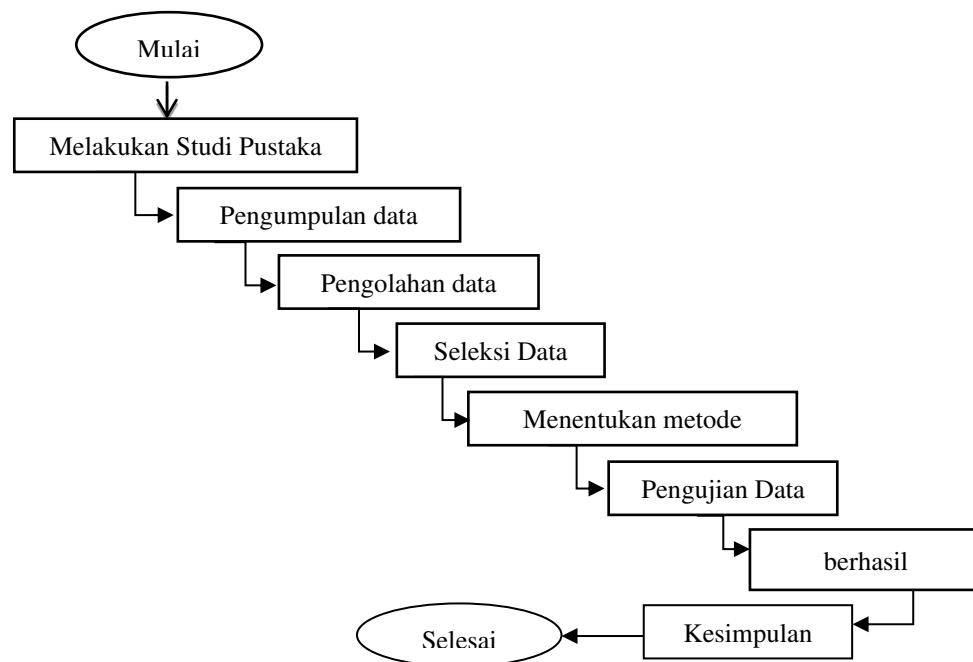
1. Tentukan Jumlah *Kluster* *k*.
2. Menentukan Titik pusat awal dari setiap *kluster*
3. Alokasikan semua data/objek ke *kluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *kluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *kluster*. Jarak paling antara satu data dengan satu *kluster* tertentu akan menentukan suatu data masuk dalam *kluster* mana. Untuk menghitung jarak semua data ke setiap titik pusat *kluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:
4. $D(x,y) = \sqrt{(X^1 - Y^1)^2 + (X^2 - Y^2)^2}$ (1)
5. Hitung kembali pusat *kluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *kluster* tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
6. Tugaskan lagi setiap objek memakai pusat *kluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *klustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *kluster* tidak berubah lagi [8].

2.3 RapidMiner

RapidMiner merupakan perangkat lunak yang dibuat untuk memudahkan pengguna dalam menggunakan perangkat lunak ini. Hasil yang diperlihatkan *RapidMiner* pun dapat ditampilkan secara *visual* dengan *grafik*, Menjadikan *RapidMiner* adalah salah satu *software* pilihan untuk melakukan ekstraksi data dengan metode-metode *data mining* [9].

2.4 Diagram Aktivitas Kerja Penelitian

Berikut ini alur diagram aktivitas kerja penelitian yang akan dilakukan pada penelitian ini :



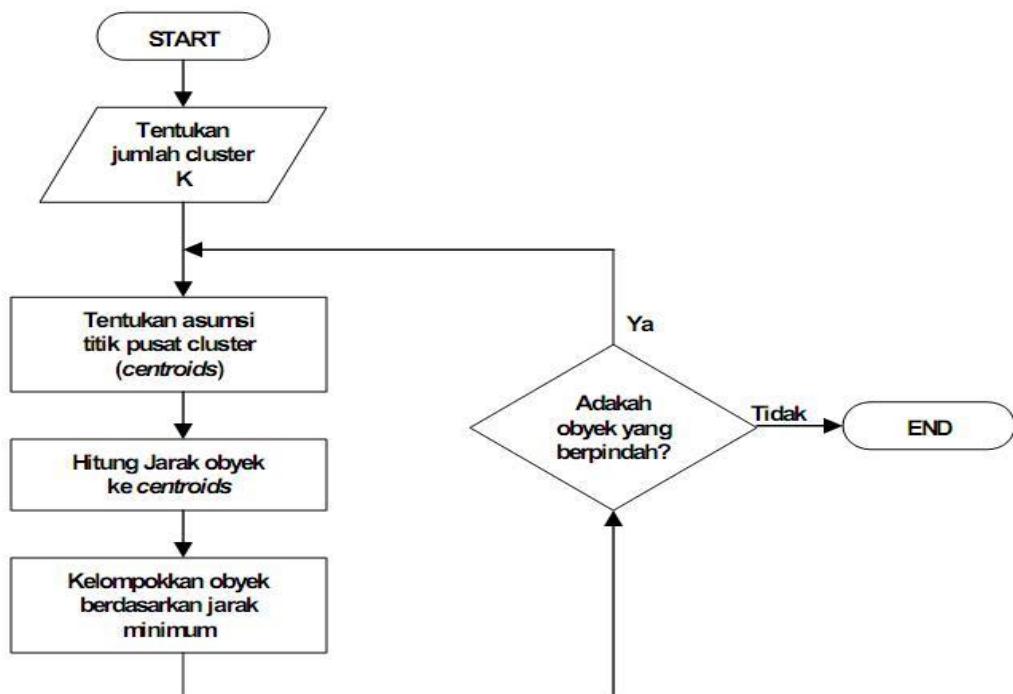
Gambar 1. Aktivitas Kerja Penelitian

Pada gambar 1 diatas dapat dijelaskan dengan beberapa urutan yaitu:

1. Pertama hal yang dilakukan adalah melakukan studi pustaka yang berkaitan dengan penelitian tentang penggunaan metode *K-means*, serta untuk menguatkan penelitian ini berdasarkan teori yang digunakan.
2. Setelah melakukan studi pustaka tahap selanjutnya adalah menentukan metode yang sesuai untuk digunakan dalam teknik pengelompokan (*cluster*), setelah melakukan pengamatan dan observasi penelitian memilih metode *K-means*.
3. Pada tahap ini peneliti melakukan pengumpulan data dan ini merupakan tahapan yang penting karena dapat berpengaruh terhadap hasil penelitian, sehingga dalam mengumpulkan data harus dilakukan dengan benar.
4. Tahap ini dilakukan seleksi terhadap data telah didapat. Karena data yang diperoleh semuanya digunakan, dan sesuai dengan atribut atau variabel yang dibutuhkan dalam penelitian dengan melakukan pengelompokan data sehingga menjadi *dataset*.
5. Setelah semua data yang diperlukan telah dipilih, maka tahap penelitian selanjutnya adalah pengolahan data. Pada tahap ini akan dilakukan transformasi atau mengubah nilai atribut data ke dalam bentuk data yang sesuai agar data dapat diproses menggunakan metode *K-means*.
6. Pada tahap pengujian hasil akan dilakukan pengujian data baik secara manual dengan metode *K-means* dan menggunakan *software rapidminer*.
7. Berdasarkan hasil pengujian maka dapat ditarik kesimpulan yang mengacu pada rumusan masalah dan tujuan penelitian. Saran yang digunakan untuk mengembangkan penelitian selanjutnya serta dimasukkan untuk meningkatkan kualitas penelitian.

2.5 Pemodelan Metode K-Means

Metode *K-means* merupakan algoritma yang relatif sederhana untuk mengklasifikasikan atau mengelompokkan sejumlah besar objek dengan atribut tertentu kedalam kelompok-kelompok sebanyak *K*. *K-means* salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada di dalam bentuk satu atau lebih *cluster* atau kelompok.



Gambar 2. Flowchart Metode K-Means

3. HASIL DAN ANALISIS

Hasil dan pembahasan pada bagian ini disajikan sesuai penelitian yang dilakukan. Pada penelitian ini digunakan *Tools Rapidminer 5.3*. sebagai pengujian data yang di analisa. Berikut merupakan penjelasan Implementasi *k-means* pada data balita yang mengalami gizi buruk usia 0-59 bulan.

3.1 Analisis Data Manual K-Means

Data penelitian terdiri dari 34 provinsi yang terdiri dari tahun 2016-2018 seperti yang tertera pada tabel 3.1. Selanjutnya masuk dalam tahap perhitungan menggunakan metode *k-means*:

1. Penentuan Pusat *Cluster* Awal

Penentuan titik *cluster* ini dilakukan dengan mengambil nilai terbesar (maksimum) untuk *cluster* tertinggi (C1), nilai terkecil untuk *cluster* rendah (C2) sebagai berikut:

Cluster	2016	2017	2018
C1	6,91	7,40	7,40
C2	0,98	2,00	2,00

2. Perhitungan Jarak *Cluster*

Untuk menghitung jarak antara data dengan pusat *cluster* memakai persamaan (1) :

$$C_{\text{aceh}}, C_1 = \sqrt{(2,55 - 6,91)^2 + (5,90 - 7,40)^2 + (6,70 - 7,40)^2} = 5,081004$$

$$C_{\text{Sumatera Utara}}, C_1 = \sqrt{(3,07 - 6,91)^2 + (5,30 - 7,40)^2 + (5,40 - 7,40)^2} = 5,043709$$

$$C_{\text{Sumatera Barat}}, C_1 = \sqrt{(2,13 - 6,91)^2 + (3,30 - 7,40)^2 + (3,50 - 7,40)^2} = 7,4974$$

$$C_{\text{Riau}}, C_1 = \sqrt{(1,96 - 6,91)^2 + (4,20 - 7,40)^2 + (4,30 - 7,40)^2} = 6,829766$$

$$C_{\text{Jambi}}, C_1 = \sqrt{(2,38 - 6,91)^2 + (3,00 - 7,40)^2 + (3,80 - 7,40)^2} = 7,364272$$

Dan untuk *cluster* C2 adalah :

$$C_{\text{aceh}}, C_2 = \sqrt{(2,55 - 0,98)^2 + (5,90 - 2,00)^2 + (6,70 - 2,00)^2} = 6,944847$$

$$C_{\text{Sumatera Utara}}, C_2 = \sqrt{(3,07 - 0,98)^2 + (5,30 - 2,00)^2 + (5,40 - 2,00)^2} = 5,618834$$

$$C_{\text{Sumatera Barat}}, C_2 = \sqrt{(2,13 - 0,98)^2 + (3,30 - 2,00)^2 + (3,50 - 2,00)^2} = 2,838538$$

$$C_{\text{Riau}}, C_2 = \sqrt{(1,96 - 0,98)^2 + (4,20 - 2,00)^2 + (4,30 - 2,00)^2} = 3,982964$$

$$C_{\text{Jambi}}, C_2 = \sqrt{(2,38 - 0,98)^2 + (3,00 - 2,00)^2 + (3,80 - 2,00)^2} = 3,01609$$

Berikut ini adalah hasil dari perhitungan jarak *cluster*:

Tabel 3. Hasil Iterasi 1

PROVINSI	TINGGI	RENDAH	JARAK	C1	C2
Aceh	5,081004	6,944847	5,081004	1	
Sumatera Utara	5,043709	5,618834	5,043709	1	
Sumatera Barat	7,4974	2,838538	2,838538		1
Riau	6,829766	3,982964	3,982964		1
Jambi	7,364272	3,01609	3,01609		1
Sumatera Selatan	7,877246	3,9219	3,9219		1
Bengkulu	8,950536	1,972917	1,972917		1
Lampung	7,938451	2,623223	2,623223		1
Kep. Bangka Belitung	7,420391	2,957448	2,957448		1
Kep. Riau	6,825907	2,986855	2,986855		1
Dki Jakarta	7,749852	1,929378	1,929378		1
Jawa Barat	8,020511	1,885975	1,885975		1
Jawa Tengah	7,308386	2,54063	2,54063		1
Di Yogyakarta	8,51693	1,574929	1,574929		1
Jawa Timur	7,042989	2,835842	2,835842		1
Banten	5,701026	3,985223	3,985223		1
Bali	9,720314	1,442498	1,442498		1
Nusa Tenggara Barat	5,460266	5,526011	5,460266	1	
Nusa Tenggara Timur	0,626259	9,654559	0,626259	1	
Kalimantan Barat	2,065672	7,737396	2,065672	1	
Kalimantan Tengah	2,647754	7,053574	2,647754	1	
Kalimantan Selatan	4,558969	5,617695	4,558969	1	
Kalimantan Timur	5,995081	3,717701	3,717701		1
Kalimantan Utara	6,311814	3,542781	3,542781		1
Sulawesi Utara	7,81796	3,538474	3,538474		1
Sulawesi Tengah	3,413517	6,421246	3,413517	1	
Sulawesi Selatan	4,170635	5,51856	4,170635	1	
Sulawesi Tenggara	5,581093	6,449566	5,581093	1	

PROVINSI	TINGGI	RENDAH	JARAK	C1	C2
Gorontalo	3,204138	7,498186	3,204138	1	
Sulawesi Barat	3,491661	6,778938	3,491661	1	
Maluku	2,181765	8,46501	2,181765	1	
Maluku Utara	6,630694	5,078819	5,078819		1
Papua Barat	2,661672	7,158128	2,661672	1	
Papua	4,630821	6,434812	4,630821	1	

Berdasarkan matrik yang didapatkan pada tabel di atas maka didapatkan pengelompokan sebagai berikut :

C1 = 1,2,18,19,20,21,22,26,27,28,29,30,31,33,34

C2 = 3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,23,24,25,32

3. Penentuan Pusat Cluster Baru

Setelah didapatkan hasil dari setiap *cluster* kemudian pusat *cluster* baru dihitung berdasarkan data member tiap-tiap *cluster* yang sudah didapatkan memakai rumus yang sesuai dengan pusat member *cluster*. Berikut adalah pusat cluster untuk iterasi 2:

Tabel 4. Pusat *Cluster* Baru

Cluster	2016	2017	2018
C1	4,5573	5,8466	5,8133
C2	2,4752	3,2421	3,3578

Perhitungan Iterasi 2:

$$C_{aceh}, C_1 = \sqrt{(2,55 - 4,5573)^2 + (5,90 - 5,8466)^2 + (6,70 - 5,8133)^2} = 3,904255$$

$$C_{SumateraUtara}, C_1 = \sqrt{(3,07 - 4,5573)^2 + (5,30 - 5,8466)^2 + (5,40 - 5,8133)^2} = 2,92144$$

$$C_{SumateraBarat}, C_1 = \sqrt{(2,13 - 4,5573)^2 + (3,30 - 5,8466)^2 + (3,50 - 5,8133)^2} = 4,601131$$

$$C_{Riau}, C_1 = \sqrt{(1,96 - 4,5573)^2 + (4,20 - 5,8466)^2 + (4,30 - 5,8133)^2} = 4,198322$$

$$C_{Jambi}, C_1 = \sqrt{(2,38 - 4,5573)^2 + (3,00 - 5,8466)^2 + (3,80 - 5,8133)^2} = 4,523808$$

Dan untuk cluster C2 adalah:

$$C_{aceh}, C_2 = \sqrt{(2,55 - 2,4752)^2 + (5,90 - 3,2421)^2 + (6,70 - 3,3578)^2} = 5,056234$$

$$C_{SumateraUtara}, C_2 = \sqrt{(3,07 - 2,4752)^2 + (5,30 - 3,2421)^2 + (5,40 - 3,3578)^2} = 3,587743$$

$$C_{SumateraBarat}, C_2 = \sqrt{(2,13 - 2,4752)^2 + (3,30 - 3,2421)^2 + (3,50 - 3,3578)^2} = 1,600363$$

$$C_{Riau}, C_2 = \sqrt{(1,96 - 2,4752)^2 + (4,20 - 3,2421)^2 + (4,30 - 3,3578)^2} = 2,490248$$

$$C_{Jambi}, C_2 = \sqrt{(2,38 - 2,4752)^2 + (3,00 - 3,2421)^2 + (3,80 - 3,3578)^2} = 1,664278$$

Berikut adalah hasil dari perhitungan jarak cluster iterasi 2:

Tabel 5. Hasil Perhitungan Iterasi 2

PROVINSI	TINGGI	RENDAH	JARAK	C1	C2
Aceh	3,904255	5,056234	3,904255	1	
Sumatera Utara	2,92144	3,587743	2,92144	1	
Sumatera Barat	4,601131	1,600363	1,600363		1
Riau	4,198322	2,490248	2,490248		1
Jambi	4,523808	1,664278	1,664278		1
Sumatera Selatan	5,406959	3,036879	3,036879		1
Bengkulu	6,003103	2,28647	2,28647		1
Lampung	5,013077	1,855378	1,855378		1
Kep. Bangka Belitung	4,518068	1,697535	1,697535		1
Kep. Riau	3,779701	0,849895	0,849895		1
Dki Jakarta	4,601117	0,474003	0,474003		1
Jawa Barat	4,950624	1,071656	1,071656		1
Jawa Tengah	4,2726	0,76923	0,76923		1
Di Yogyakarta	5,468178	1,513953	1,513953		1
Jawa Timur	4,035167	0,892878	0,892878		1
Banten	2,608073	1,617884	1,617884		1
Bali	6,68565	2,724534	2,724534		1
Nusa Tenggara Barat	3,463761	3,600297	3,463761	1	
Nusa Tenggara Timur	3,337326	7,29271	3,337326	1	

PROVINSI	TINGGI	RENDAH	JARAK	C1	C2
Kalimantan Barat	1,254489	5,385918	1,254489	1	
Kalimantan Tengah	0,968922	4,69135	0,968922	1	
Kalimantan Selatan	2,326516	3,395422	2,326516	1	
Kalimantan Timur	2,847878	1,426813	1,426813		1
Kalimantan Utara	3,083311	1,447431	1,447431		1
Sulawesi Utara	5,177185	2,670409	2,670409		1
Sulawesi Tengah	0,919371	4,104882	0,919371	1	
Sulawesi Selatan	1,255339	3,150697	1,255339	1	
Sulawesi Tenggara	3,959378	4,699835	3,959378	1	
Gorontalo	2,59341	5,25926	2,59341	1	
Sulawesi Barat	2,12864	4,495187	2,12864	1	
Maluku	2,849171	6,146265	2,849171	1	
Maluku Utara	4,499427	3,595414	3,595414		1
Papua Barat	0,948354	4,825124	0,948354	1	
Papua	2,852679	4,424873	2,852679	1	

Berdasarkan matrik yang didapatkan pada tabel di atas maka didapatkan pengelompokan sebagai berikut:

C1 = 1,2,18,19,20,21,22,26,27,28,29,30,31,33,34

C2 = 3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,23,24,25,32

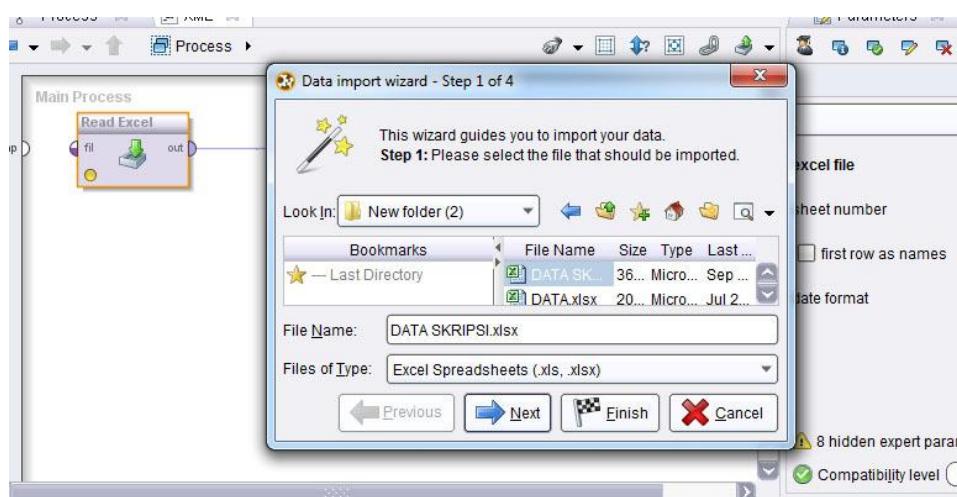
Dari hasil iterasi 2 dapat dilihat bahwa hasil yang diperoleh sama dengan iterasi 1 maka perhitungan dihentikan sampai iterasi 2. Untuk provinai pada cluster C1(tertinggi) ada 15 Provinsi yaitu (Aceh, Sumatera Utara, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Papua Barat, Papua). Untuk provinsi pada cluster C2(rendah) ada 19 yaitu (Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Dki Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Timur, Kalimantan Utara, Sulawesi utara, Maluku Utara).

3.2 Pemodelan RapidMiner

Pada penelitian ini digunakan *tools rapidminer* sebagai pengujian data yang telah di analisa. Berikut adalah bagian-bagian dari *rapidminer* beserta isi dari data yang dibuat.

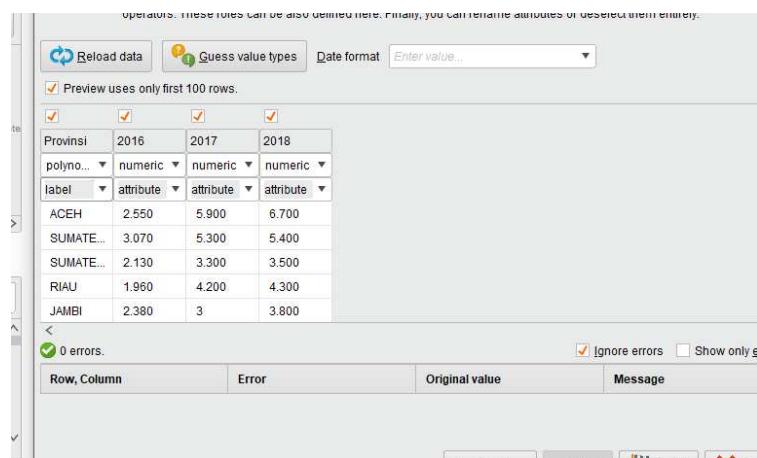
1. Input Sistem

Berikut adalah penjelasan cara memasukkan data baru yang akan dieksekusi lebih lanjut, pada hal ini data yang akan di eksekusi berupa data *excel*. Tahapan dapat dilihat pada gambar dibawah ini.



Gambar 3. Process Import Data Excel

Untuk menginputkan data *excel* dapat menggunakan dua cara yaitu melalui bagian *filter* dengan mengetikkan *read excel* atau dapat melalui bagian *repositories* lalu *Import Excel Sheet*. Pada bagian look in kita dapat mencari dimana letak file data *excel* yang kita simpan. Seperti pada gambar, data disimpan dalam folder data dengan nama data seperti yang tertera pada gambar. Selanjutnya akan muncul tampilan seperti gambar 4 :



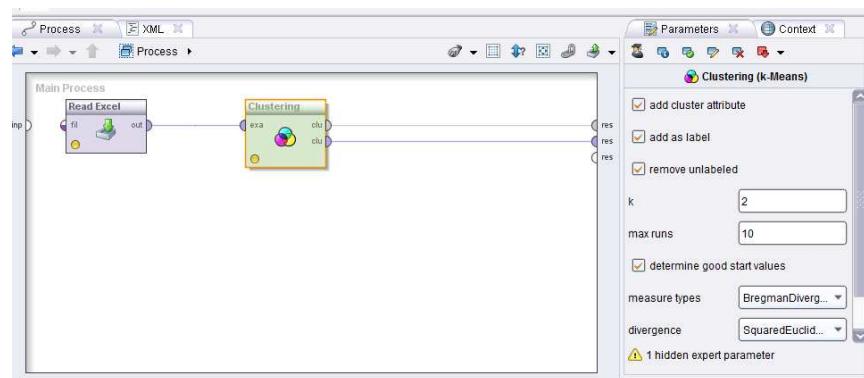
Provinsi	2016	2017	2018
polyno...	numeric	numeric	numeric
label	attribute	attribute	attribute
ACEH	2.550	5.900	6.700
SUMATE...	3.070	5.300	5.400
SUMATE...	2.130	3.300	3.500
RIAU	1.960	4.200	4.300
JAMBI	2.380	3	3.800

Gambar 4. Import Data Excel

Pada ini dilakukan pemilihan tipe data dimana pada tahap ini memilih bagian yang akan diberi tipe “label” yang akan menjadi penentu dari dibentuknya pengelompokan.

2. Proses Sistem

Lakukan *drag and drop* data yang dipilih kedalam *view process*. Berikut alur proses *import* data yang dilakukan.



Gambar 5. Proses Sistem

Setelah data yang dipilih didrag kedalam *main process*, selanjutnya pada bagian *filter* ketikkan *k-means*. Maka akan muncul operator *k-means* lalu pilih dan drag kedalam halaman *main process*. Selanjutnya hubungkan data *read excel* dengan operator *k-means* lalu klik tools run.

3. Output Sistem

Setelah melakukan proses sistem seperti yang tertera pada gambar 5 maka akan muncul hasil dari *rapidminer* yang dapat dilihat pada gambar 6, dimana *cluster 0* adalah *cluster* tinggi dan *cluster 1* adalah *cluster* rendah.

Row No.	id	Provinsi	cluster	2016	2017	2018
1	1	ACEH	cluster_0	2.550	5.900	6.700
2	2	SUMATERA ...	cluster_0	3.070	5.300	5.400
3	3	SUMATERA B...	cluster_1	2.130	3.300	3.500
4	4	RIAU	cluster_1	1.960	4.200	4.300
5	5	JAMBI	cluster_1	2.380	3	3.800
6	6	SUMATERA S...	cluster_1	1.930	2.100	4.900
7	7	BENGKULU	cluster_1	1.300	2.300	2.800
8	8	LAMPUNG	cluster_1	1.630	3.500	3.100
9	9	KEP. BANGK...	cluster_1	2.010	3.700	3.400
10	10	KEP. RIAU	cluster_1	3.730	3	3.200
11	11	DKI JAKARTA	cluster_1	2.990	3	2.300
12	12	JAWA BARAT	cluster_1	2.350	2.900	2.600
13	13	JAWA TENGAH	cluster_1	2.980	3	3.100
14	14	DI YOGYAKA...	cluster_1	2.100	2.400	2.500
15	15	JAWA TIMUR	cluster_1	3.360	2.900	3.300

Gambar 6. Data Hasil Clustering

4. KESIMPULAN

Berdasarkan pembahasan sebelumnya dapat disimpulkan bahwa penerapan *Data mining* dengan menggunakan algoritma k-means pada pengelompokan balita yang memiliki gizi buruk menurut provinsi dapat diterapkan. Sumber data yang digunakan pada penelitian ini adalah data yang diperoleh dari BPS (Badan Pusat Statistik). Jumlah data yang digunakan sebanyak 34 provinsi yang terdiri dari tahun 2016-2018. Dari hasil pengelompokan diperoleh dua cluster yaitu tinggi dan renda. Cluster tinggi terdiri dari 15 provinsi dan cluster rendah terdiri dari 19 provinsi. Pengujian data pada *Rapidminer 5.3* menggunakan algoritma k-means dapat menampilkan keakuratan data antara perhitungan manual dan sistem.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih pada STIKOM Tunas Bangsa Pematangsiantar dan pada pihak-pihak yang membantu penulis dalam menyelesaikan penelitian ini.

REFERENSI

- [1] S. D. Nugraha, R. R. M. Putri, and R. C. Wihandika, “Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita,” vol. 1, no. 9, pp. 925–932, 2017.
- [2] A. N. Khomarudin, “Teknik Data Mining : Algoritma K-Means Clustering,” pp. 1–12, 2016.
- [3] M. R. Ridlo, S. Defiyanti, and A. Primajaya, “Implementasi Algoritme K-Means Untuk Pemetaan Produktivitas Panen Padi Di Kabupaten Karawang,” pp. 426–433, 2017.
- [4] R. Wahyuni, S. Nugroho, and P. Novianti, “Analisis Klaster Dengan Menggunakan Metode Single Linkage Dan Metode K-Means,” pp. 1–9, 2015.
- [5] Y. Warih, Eggy Inaidi Andana; Rahayu, “Penerapan Data Mining untuk Menentukan Estimasi Produktivitas Tanaman Tebu dengan Menggunakan Algoritma Linear Regresi Berganda di Kabupaten Rembang,” *Informatika*, pp. 1–5, 2014.
- [6] M. R. Ridlo, S. Defiyanti, A. Primajaya, M. Rosyid Ridlo, S. Defiyanti, and A. Primajaya, “Implementasi Algoritme K-Means Untuk Pemetaan Produktivitas Panen Padi Di Kabupaten Karawang,” *Citee 2017*, pp. 426–433, 2017.
- [7] E. Rivani, P. Pengkajian, P. Data, S. Jenderal, and D. P. R. Ri, “APLIKASI K- MEANS CLUSTER UNTUK PENGELOMPOKKAN PROVINSI BERDASARKAN PRODUKSI PADI , JAGUNG , KEDELAI , DAN KACANG HIJAU TAHUN 2009 Ukuran Similaritas,” vol. 10, no. 2, pp. 122–134, 2010.
- [8] D. Triyansyah and D. Fitrianah, “Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing,” no. September, 2018.
- [9] T. Verma, “Tokenization and Filtering Process in RapidMiner,” vol. 7, no. 2, pp. 16–18, 2014.