Penentuan Kategori Status Gizi Balita Menggunakan Penggabungan Metode Klasterisasi *Agglomerative* Dan *K-Means*

Aries Aprilia[1], Weny Mistarika Rahmawati[2], dan Maftahatul Hakimah[3]
Institut Teknologi Adhi Tama Surabaya[1,2,3]
e-mail: ariesaprilia415@gmail.com[1]

ABSTRACT

Data clustering is widely used in various fields. One of them is used for grouping the nutritional status of toddlers. The clustering method that is often used is the k-means method, which makes the nearest neighbor as a cluster. The disadvantage of k-means is the placement of random initial centroids. Laying different initial centroids can produce different cluster results. Therefore, in this study the grouping of the nutritional status of toddlers using the k-means method with centroid placement was initially obtained from the agglomerative clustering method. At first the data is processed by agglomerative clustering method to produce the desired number of clusters. The centroid of each cluster result will be used as the initial centroid of the k-means algorithm. In this way, the disadvantages of k-means by finding the initial centroid at random can be overcome. Based on testing the distance between centroids, the combined agglomerative and k-means clustering methods are better than only k-means. Meanwhile, the elbow graph shows that the optimal number of clusters for grouping the nutritional status of toddlers is 3 clusters.

Keyword: *Clustering, agglomerative, k-means.*

ABSTRAK

Klasterisasi data banyak digunakan dalam berbagai bidang. Salah satunya digunakan untuk pengelompokan status gizi balita. Metode klasterisasi yang sering digunakan adalah metode *k-means*, yaitu menjadikan tetangga terdekat sebagai satu klaster. Kelemahan dari *k-means* adalah peletakan titik pusat klaster (*centroid*) awal yang acak. Peletakan *centroid* awal yang berbeda dapat menghasilkan hasil klaster yang berbeda. Oleh karena itu, pada penelitian ini pengelompokan status gizi balita menggunakan metode *k-means* dengan peletakan *centroid* awalnya diperoleh dari metode klasterisasi *agglomerative*. Mula-mula data diolah dengan metode klasterisasi *agglomerative* menghasilkan klaster sejumlah yang diinginkan. Centroid dari setiap hasil klaster tersebut akan dijadikan sebagai *centroid* awal algoritma k-means. Dengan cara ini, kelemahan k-means dengan mencari *centroid* awal secara acak dapat diatasi. Berdasarkan pengujian jarak antar *centroid*, gabungan metode klasterisasi *agglomerative* dan *k-means* lebih baik daripada *k-means* saja. Sedangkan, grafik *elbow* menunjukkan bahwa jumlah klaster yang optimal untuk pengelompokan status gizi balita adalah sebanyak 3 klaster.

Kata kunci: Klasterisasi, agglomerative, *k-means*.

PENDAHULUAN

Permasalahan pemenuhan gizi balita selalu menjadi perhatian pemerintah Salah satu bentuk program pemerintah untuk mengontrol pertumbuhan balita adalah pos pelayanan terpadu (posyandu). Kegiatan rutin setiap bulan oleh posyandu ini dilakukan pada balita untuk mengetahui status gizinya. Status gizi balita di posyandu ini umumnya dikelompokkan menjadi 5 kategori yaitu gizi buruk, gizi kurang, gizi normal, gizi lebih dan obesitas. Namun, karena petugas posyandu yang terbatas dengan banyak tugas yang harus dikerjakan misalnya, mencatat rekam medis balita manual, pemberian imunisasi, penyuluhan kepada masayarakat, dll maka penanganan terhadap status gizi balita menjadi terhambat. Penanganan terhadap status gizi merupakan salah satu strategi dari pemerintah dalam menciptakan sumber daya manusia yang sehat, cerdas dan produktif. Upaya ini dilakukan dengan menangani pertumbuhan anak secara dini dengan memperhatikan asupan gizi dan perawatan yang baik [9].

Dari permasalahan yang terjadi di kebanyakan posyandu ini, peneliti menerapkan metode klasterisasi untuk mengelompokkan status gizi balita dengan bantuan komputer sehingga membantu petugas posyandu lebih efektif dan efisien dalam melaksanakan tugasnya. Metode klasterisasi adalah

metode pengelompokan data ke dalam beberapa *klaster* (kelompok) sedemikian hingga satu klaster berisi objek yang mempunyai karakteristik yang hampir mirip dan jauh berbeda dengan objek pada klaster yang lain. Metode klasterisasi yang digunakan penelitian ini adalah menggabungkan metode klasterisasi *agglomerative* model *single linkage* dan metode *k-means*. Model *single linkage* digunakan untuk menentukan *centroid* (titik pusat klaster) awal yang akan digunakan pada metode *k-means* dalam menentukan status gizi. Pada umumnya, centroid pada metode *k-means* ditentukan secara acak sehingga penyelesaiannya bersifat *local optimum* [1].

Penentuan status gizi balita pada penelitian ini melibatkan 3 faktor yaitu, umur, berat badan dan tinggi badan balita. Kinerja dari metode gabungan *single linkage* dan *k-means* ini dalam mengelompokkan data diukur menggunakan rumus *sum of square error* (SSE). Jumlah klaster pada penelitian ini akan diuji coba untuk beberapa *k* klaster sehingga bisa diketahui data akan mengelompok dalam klaster yang tepat. Jumlah klaster optimal ini dicari menggunakan. nilai SSE yang disajikan dalam grafik *elbow*. Jumlah klaster optimal ditentukan dari nilai SSE pada *k* klaster yang mengalami penurunan terbesar sehingga membentuk sudut pada grafik tersebut [7]. Selain itu, pengukuran kinerja penggabungan metode ini dilihat dari jarak antar titik pusat klaster yang terbentuk.

TINJAUAN PUSTAKA

Klasterisasi *Agglomerative*

Klasterisasi adalah metode penganalisaan data yang sering dimasukkan sebagai salah satu metode yang bertujuan untuk mengelompokan data dengan karakteristik yang sama ke dalam wilayah yang sama dan karakteristik yang beda ke dalam wilayah yang lain. Klasterisasi *Agglomerative* merupakan bagian dari klastering berdasarkan hirarki. Klasterisasi *Agglomerative* dimulai dengan membuat setiap data merupakan sebuah klaster. Klaster-klaster tersebut akan bergabung berdasarkan jarak antar klaster. Penggabungan klaster akan terus terjadi sampai klaster yang dihasilkan sejumlah yang diinginkan[2].

Untuk menghitung jarak antar klaster pada algoritma *agglomerative* dapat dilakukan menggunakan beberapa cara. Salah satunya menggunakan metode *single linkage*. Algoritma s*ingle linkage* adalah mencari jarak dua klaster sesuai dengan jarak terpendek antara dua anggota dalam dua *klaster* [5]. Pada metode ini kedekatan diantara dua kelompok ditentukan dari jarak terdekat antara pasangan dua data dari dua kelompok yang berbeda (satu dari kelompok pertama dan satu lagi dari kelompok yang lain). Formulasinya dapat dilihat pada Persamaan 1 berikut ini.

$$d_{uv} = \min\{d_{uv}\}, d_{uv} \in d \tag{1}$$

K-Means

K-Means adalah suatu metode pengelompokan data non-hirarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok lain [8].

Algoritma *K-Means* dapat dilakukan dengan cara sebagai berikut:

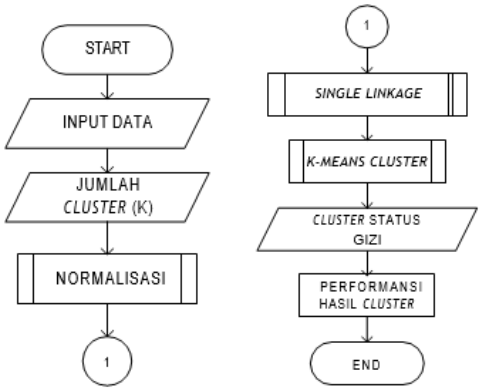
1. Tentukan *k* sebagai jumlah klaster yang diinginkan dan matriks jarak yang diinginkan.
2. Pilih *k* dari himpunan data *X* sebagai *centroid*.
3. Alokasikan semua data dari *centroid* terdekat dengan matriks jarak yang sudah ditetapkan dengan menggunakan rumus *Euclidean Distance* berikut:

$$\mathrm{Dist}(\overline{X_i}, \overline{Y_j}) = \left\| \overline{X_i}, \overline{Y_j} \right\|^2 \tag{2}$$

4. Hitung kembali *centroid* baru berdasarkan data yang mengikuti klaster masing-masing.
5. Ulangi langkah (3) dan (4) hingga kondisi tidak ada data yang berpindah.

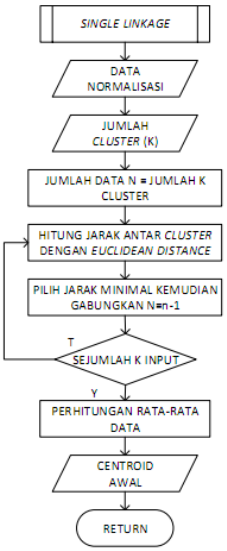
METODE

Flowchart metode klasterisasi yang digunakan pada penelitian ini dapat dilihat pada Gambar 1.



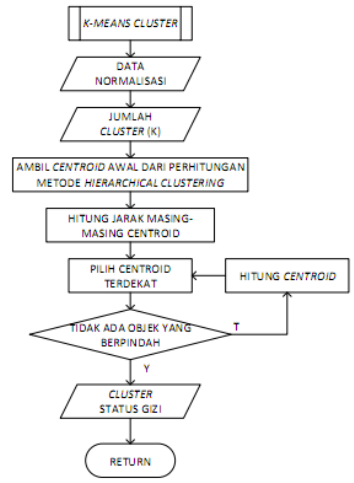
Gambar 1 Flowchart Penentuan Status Gizi Balita

Gambar 1 menunjukkan langkah-langkah untuk menentukan status gizi balita yaitu input data meliputi data umur, berat badan dan tinggi badan yang sebelumnya dibersihkan dari data yang sama (*redundant*) dengan memakai aplikasi yang *extends* dengan *Ms. Excel* bernama *Kutools*. Selanjutnya tentukan jumlah *klaster* yang bisa dipilih oleh user. Normalisasi data diperlukan untuk menyamakan satuan dimana umur dengan satuan bulan, berat badan dengan satuan kg, dan tinggi badan dengan satuan cm. Setelah mendapatkan nilai normalisasi, selanjutnya adalah proses perhitungan dengan menggunakan metode *single linkage*. Langkah berikutnya adalah menentukan centroid awal untuk metode *k-means* menggunakan klaterisasi *agglomerative model single linkage*. Berikut ini merupakan tahapan untuk metode *single linkage*.



Gambar 2 Flowchart *Agglomerative* dengan *single linkage*

Gambar 2 merupakan *flowchart* proses metode *single linkage*. Dari data yang ternormalisasi dan input jumlah k, berikutnya hitung jarak antar cluster dengan *Euclidean distance* dan pilih jarak minimal. Iterasi berjalan sampai *k* tercapai. Setelah iterasi berhenti, dilakukan perhitungan rata-rata dari jumlah data pada iterasi terakhir. Nilai rata-rata ini merupakan *centroid* awal yang akan digunakan pada metode *k-means* terbentuk. Berikut ini adalah tahapan metode *k-means.*



Gambar 3 Flowchart metode *k-means*

Gambar 3 merupakan *flowchart* dari metode *k-means*. *Centroid* yang dihasilkan oleh metode *agglomerative* (*hierarchical clustering*) digunakan sebagai inisialisasi *centroid* pada metode *k-means*. Selanjutnya, hitung jarak masing-masing centroid. Dari perhitungan jarak ini, centroid akan dicari secara iteratif sampai tidak ada objek yang berpindah dalam klaster yang dibentuk. Jika kondisi ini dipenuhi maka metode *k-means* menghasilkan klaster-klaster status gizi sesuai dengan jumlah k yang diinputkan.

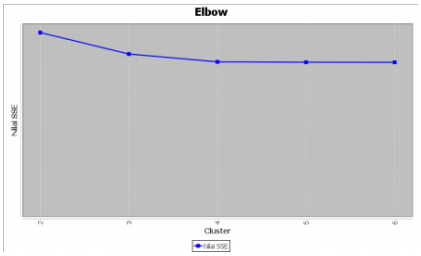
HASIL DAN PEMBAHASAN

Pengujian sistem pada klasterisasi menggunakan penggabungan metode klasterisasi *Agglomerative* model *single linkage* dan *k-means* dilakukan pada 2849 data balita yang ada di posyandu Pinggir Papas Madura. Data diuji dengan mengklasterkan secara agglomerative yang dilanjutkan dengan *k- means*. Selanjutnya sebagai perbandingan, data yang sama diklasterkan menggunakan k-means saja tanpa klasterisasi *agglomerative*. Pengujian dilakukan dengan menghitung SSE didalam klaster dan juga menghitung jarak antar klaster. Selain itu, pengujian dilakukan dengan mengklasterkan data menjadi 2 klaster, 3 klaster, 4 klaster, 5 klaster dan 6 klaster. Tabel 1 berikut ini merupakan hasil pengujian yang diperoleh.

Tabel 1 Hasil pengujian Klaster

	SSE		**JARAK ANTAR *CENTROID***	
Klaster (k)	Aglo+Kmeans	Kmeans	Aglo+Kmeans	Kmeans
k = 2	50.30589	15.76968	0.62623	0.43593
k = 3	44.42754	9.72576	0.74069	0.39213
k = 4	42.28758	7.89193	0.79911	0.37961
k = 5	42.18364	4.35537	0.77545	0.51791
k = 6	42.14903	6.67621	0.78404	0.34871

Hasil pengujian pada Tabel 1 menunjukkan bahwa semakin banyak klaster yang dibentuk maka SSE akan semakin rendah. Berdasarkan nilai SSE, metode k-means lebih baik daripada metode gabungan klasterisasi *agglomerative* dan *k-means*. Namun, jika dilihat pada jarak antar *centroid*, metode gabungan klasterisasi *agglomerative* dan *k-means* menunjukkan hasil klasternya lebih baik daripada *k-means*. Selanjutnya, metode elbow diterapkan untuk mengetahui jumlah klaster terbaik untuk mengelompokkan data status gizi balita.



Gambar 4 Grafik Elbow untuk k = 2, 3, 4, 5, 6

Dari Gambar 4 dapat dilihat bahwa penurunan nilai SSE terbesar terjadi pada k=3 sehingga membentuk sudut pada grafik tersebut. Oleh karena itu, jumlah klaster optimal untuk mengelompokkan status gizi balita adalah 3 klaster.

KESIMPULAN

Pada penelitian ini melakukan klasterisasi gizi balita menggunakan algoritma gabungan klasterisasi *agglomerative* dan algoritma k-means. Hasil dari *Klasterisasi Agglomerative* berupa *centroid* yang digunakan untuk *centroid* algoritma k means. Hasil dari penelitian ini dapat disimpulkan bahwa jumlah klaster yang optimal untuk mengklasterkan status gizi balita adalah sebanyak 3 klaster dengan nilai SSE 44,42754 dan jarak antar klaster 0,74069.

DAFTAR PUSTAKA

[1] T. Alfina, B. Santosa, and J. A. R. Hakim, "Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)," vol. 1, p. 5, 2012.

[2] C. C. Anggarwal, *Data Clustering Alorithm and Applications*, vol. 20121229. Chapman and Hall/CRC, 2012.

[3] G. Gan, C. Ma, and J. Wu, *Data clustering: theory, algorithms, and applications*. Philadelphia, Pa. : Alexandria, Va: SIAM, Society for Industrial and Applied Mathematics ; American Statistical Association, 2007.

[4] N. P. E. Merliana, Ernawati, and Alb. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering," pp. 1–7.

[5] E. Mooi and M. Sarstedt, *A concise guide to market research: the process, data, and methods using IBM SPSS statistics*. Berlin: Springer, 2011.

[6] A. F. Muhammad, "KLASTERISASI PROSES SELEKSI PEMAIN MENGGUNAKAN ALGORITMA K-MEANS," p. 5.

[7] E. Muningsih, "Optimasi Jumlah Cluster K-Means Dengan Metode Elbow Untuk Pemetaan Pelanggan," pp. 105–114, Sep. 2017.

[8] E. Prasetyo, *Data Mining - Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi Offset, 2012.

[9] P. Sari and B. Pramono, "IMPROVE K-MEANS TERHADAP STATUS NILAI GIZI PADA BALITA," p. 6.

[10] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 336, p. 012017, Apr. 2018.

ISSN (print): 2686-0023
ISSN (online): 2685-6875

Halaman ini sengaja dikosongkan