

OPTIMALISASI ALGORITMA K-MEANS MENGGUNAKAN METODE PSO PADA PENYAKIT STUNTING

Ulumuddin

Universitas Bina Sarana Informatika
Email: 1ulumuddin.udn@bsi.ac.id

Abstrak

Stunting merupakan gangguan pertumbuhan linier pada bayi baru lahir yang di sebabkan oleh beberapa factor yaitu antara lain factor BBLR berat badan lahir rendah, pendidikan ibu, pendapatan rumah tangga, dan lain sebagainya. Sehingga diperlukannya pengklasteran pada kasus ini agar mengetahui stunting dengan cluster ringan, berat ataupun sedang. Pada penelitian ini peneliti menggunakan algoritma k-means dalam melakukan pengklasteran. Sehingga mendapatkan hasil cluster berat sebanyak 41, cluster sedang sebanyak 109 sedangkan cluster ringan sebanyak 50. Agar mengetahui tingkat akurasi pada algoritma k-means yang dioptimasikan dengan PSO, dari hasil uji coba yang peneliti lakukan untuk optimasi k-means PSO, PSO terbukti bisa meningkatkan nilai akurasi terhadap k-means standar dengan nilai akurasi 78,88%

Kata kunci: *algoritma k-means, pso, stunting*

Abstract

Stunting is a linear growth disorder in newborn babies which is caused by several factors, including LBW factors, low birth weight, mother's education, household income, and so on. So clustering is needed in this case to identify stunting with mild, severe or moderate clusters. In this study, researchers used the k-means algorithm to carry out clustering. So the results obtained were 41 heavy clusters, 109 medium clusters, while 50 light clusters. In order to find out the level of accuracy in the k-means algorithm optimized with PSO, from the results of trials conducted by researchers to optimize PSO k-means, PSO was proven to be able to increase the accuracy value of standard k-means with an accuracy value of 78.88%

Keywords: *algorithm k-means, pso, stunting*

1. PENDAHULUAN

.Stunting adalah gangguan pertumbuhan linier yang di sebabkan kurang gizi pada balita sehingga berefek pada gangguan pertumbuhan pada anak balita terutama pada pertumbuhan tinggi badan. Stunting juga menjadi penyebab utama anak – anak balita pada meninggal pada seluruh dunia hingga 2,2 juta jiwa.

Dalam kasus ini peneliti melakukan penelitian dengan mengklaster pada pasien yang terkena stunting, Dalam kasus ini peneliti menggunakan algoritma k-means untuk dilakukannya klasterisasi pada jenis kelamin pasien stunting.

Menurut Han dan Kamber secara umum metode klastering dapat diklasifikasikan menjadi empat kategori dasar yaitu metode partisi, metode hirarki, metode berbasis jarak, dan metode berbasis grid. Metode klastering yang paling populer adalah metode hirarki dan metode partisi metode klastering adalah proses pengelompokan sekumpulan objek data ke dalam beberapa kelompok atau klaster sehingga data dalam sebuah klaster memiliki kemiripan yang maksimum, dan data antar klaster memiliki tingkat kemiripan yang minimum Aplikasi pengklasteran dipakai di berbagai bidang, seperti data mining, pengenalan pola, klasifikasi citra, ilmu-ilmu yang terkait dengan biologi, pemasaran, perencanaan kota, document retrieval, dan lain-lain.

K-means merupakan sebuah metode penganalisaan data atau salah satu metode data mining yang merupakan proses permodelan tanpa supervisi (unsupervised) serta merupakan salah satu metode yang melakukan permodelan data dengan mengelompokan data naik skala kecil atau besar dengan sistem partisi.

2. METODE PENELITIAN

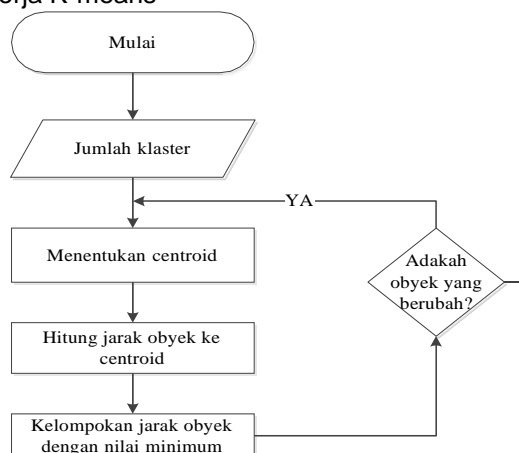
Dalam penelitian ini peneliti menggunakan metode algoritma k-means untuk mengelompokkan pasien stunting. Kemudian menggunakan metode K-Means yang dioptimalkan dengan PSO agar mendapatkan hasil akurasi maksimal pada penentuan jarak pada k-means.

2.1. Algoritma K-Means

Metode – metode algoritma clustering yang biasa digunakan ada 4 macam, yaitu metode hirarki, metode partisi, metode density dan metode grid. Sedangkan untuk algoritma K-means itu sendiri menggunakan metode partisi yang berbasis centroid. Adapun untuk penentuan nilai pada algoritma k-means itu sendiri berdasarkan nilai K partisi, penentuan titik centroid dan penentuan jarak.

K-Means merupakan salah satu algoritma clustering yang berbasis centroid dengan penentuan nilai baik pada penentuan jumlah K partisi, penentuan titik centroid ataupun penentuan nilai jarak. Algoritma ini juga menerima masukan berupa data tanpa label, hal ini berbeda dengan *supervised learning* yang menerima masukan berupa vektor $(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)$, di mana x_i merupakan data dari suatu data pelatihan dan y_i merupakan label kelas untuk x_i

Langkah – langkah untuk menentukan nilai cluster pada algoritma clustering adalah sebagai berikut, 1. Tentukan nilai K pada jumlah cluster yang akan dibentuk, 2. Tentukan nilai titik pusat cluster dengan cara dirandom/acak, 3. Hitung jarak dari masing – masing centroid dari nilai cluster, untuk menghitung jarak antara objek dengan centroid bisa menggunakan Euclidian Distance, 4. Alokasikan masing – masing objek pada nilai centroid yang paling dekat, 5. Lakukan iterasi kemudian tentukan posisi titik centroid baru dengan persamaan, 6. Ulangi langkah 3 bilamana posisi centroid baru tidak sama (Hartigan and Wong 1979), dibawah ini gambar alur kerja K-means



Gambar 1. Metode Penelitian

Pada gambar diatas menjelaskan langkah – langkah implementasi algoritma k-means. Berikut penjelasannya.

A. Tentukan banyaknya kluster, pada penelitian ini peneliti menggunakan 3 kluster ($k=3$). Banyaknya kluster harus lebih kecil dari pada banyaknya data. Adapun label yang dijadikan sebagai cluster yaitu seperti yang ada pada tabel dibawah ini,

Tabel 1. Klasterisasi

Cluster 1	BBLR
Cluster 2	PENDIDIKAN IBU
Cluster 3	PENDAPATAN RUMAH TANGGA

B. Tentukan centroid pada setiap cluster. Untuk menentukan centroid pada kluster. Banyak metode yang bisa digunakan salah satunya adalah dengan cara radom / acak yang diambil dari dataset yang

digunakan. Pada tabel dibawah ini merupakan contohnya yang digunakan sebagai titik centroid yang diambil dari 200 dataset.

Tabel 2. Titik Centroid dari 200 dataset

BBLR	PENDIDIKAN IBU	PENDAPATAM RUMAH TANGGA
66	32	117
80	35	117
66	39	119

- C. Menghitung jarak data dengan centroid, untuk menghitung jarak banyak rumus bisa yang digunakan, ada euclidean, manhatan / city block dan minkowski. Namun disini peneliti menggunakan euclidean untuk menghitung jarak entroid. Berikut rumus dan penjelasannya,

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Keterangan :

d = jarak

j = banyaknya data

c = centroid

x = data

c = centroid

jarak data dengan cluster 1, 2 dan 3 pada implementasi penelitian ini sebagai berikut,

cluster 1

$$d(x_1, C_1) = \sqrt{(66 - 80)^2 + (32 - 35)^2 + (117 - 117)^2 + (102 - 100)^2} = 0,0$$

$$d(x_1, C_1) = \sqrt{(66 - 75)^2 + (32 - 37)^2 + (117 - 117)^2 + (102 - 104)^2} = 2,6$$

$$d(x_1, C_1) = \sqrt{(66 - 70)^2 + (32 - 33)^2 + (117 - 117)^2 + (102 - 100)^2} = 5,2$$

Cluster 2

$$d(x_1, C_1) = \sqrt{(80 - 65)^2 + (35 - 30)^2 + (117 - 118)^2 + (100 - 103)^2} = 8,6$$

$$d(x_1, C_1) = \sqrt{(80 - 66)^2 + (35 - 36)^2 + (117 - 117)^2 + (100 - 100)^2} = 6,7$$

$$d(x_1, C_1) = \sqrt{(80 - 78)^2 + (35 - 35)^2 + (117 - 117)^2 + (100 - 107)^2} = 8,6$$

Cluster 3

$$d(x_1, C_1) = \sqrt{(66 - 75)^2 + (39 - 35)^2 + (118 - 117)^2 + (101 - 107)^2} = 3,3$$

$$d(x_1, C_1) = \sqrt{(66 - 69)^2 + (39 - 32)^2 + (118 - 117)^2 + (101 - 100)^2} = 8,3$$

$$d(x_1, C_1) = \sqrt{(66 - 82)^2 + (39 - 25)^2 + (118 - 117)^2 + (101 - 104)^2} = 6,$$

- D. Kelompokkan data sesuai dengan clusternya, yaitu data yang memiliki jarak terdekat, seperti pada tabel berikut,

Tabel 3. Jarak terdekat antar cluster

centroid 1			centroid 2				centroid 3				cluster 1	cluster 2	cluster 3
32	117	102	80	35	117	100	66	39	118	101			
0.00			14.46				7.14				YES		
15.00			8.60				17.72					YES	
2.65			16.12				9.27				YES		
10.49			6.71				9.75					YES	

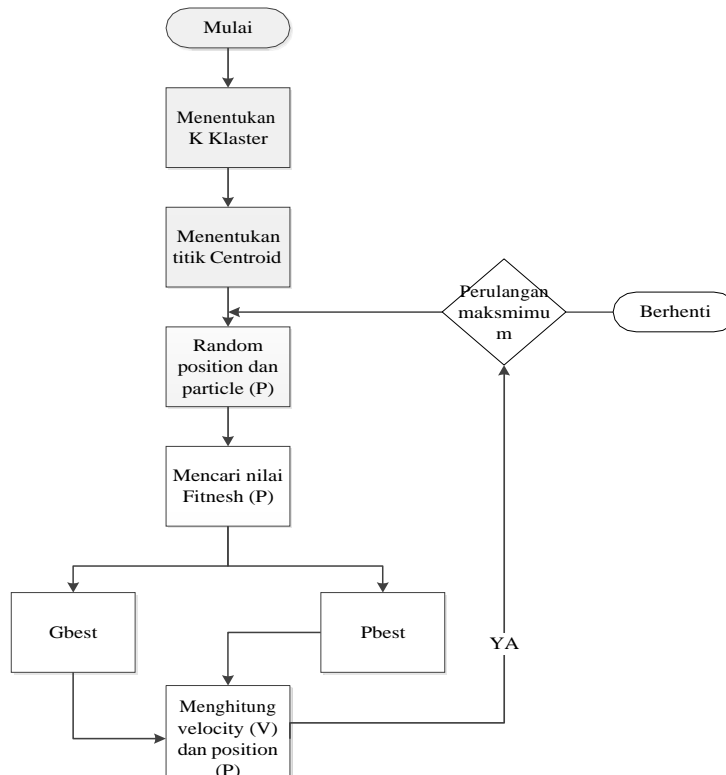
E. Jika nilai centroid pada klaster masih ada yang berubah maka ulangi langkah 3 dan 4 sampai tidak ada perubahan.

2.2. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) merupakan algoritma yang berbasis populasi yang terinspirasi dari fenomena alam perilaku burung dan ikan. Ini merupakan metode meta-heuristik yang berbasis populasi yang mengoptimalkan masalah dengan menganalisis perilaku kawanan burung atau ikan secara acak dalam ruang pencarian dimana setiap burung disebut sebagai "particle" dan *population of particles* yang disebut "swarm"

Dalam algoritma PSO kawanan disumsikan mempunyai ukuran tertentu pada setiap partikel posisi awalnya terletak pada lokasi yang acak, setiap partikel mempunyai dua karakteristik, yaitu posisi dan kecepatan, setiap partikel bergerak pada space tertentu dan mengingat posisi yang telah dilalui atau ditemukan sumber makanan atau fungsi nilai objektif. (Santosa, Implementasi, and Ilmu 2011)

2.3. K-means PSO



Gambar 2. Alur Kerja Algoritma K-Means

Pada gambar 2 menjelaskan alur kerja dari algoritma k-means yang dioptimalkan dengan PSO, berikut penjelasannya.

1. Pada penelitian ini peneliti menggunakan 3 klaster
2. Menentukan titik centroid pada klaster, yaitu dengan cara radom
3. menentukan nilai position atau particle dengan cara radom atau acak
4. Setelah melakukan langkah 1, 2 dan 3 selanjutnya adalah mencari nilai fitness dimana nilai fitness didapat dari nilai SSE. Nilai SSE merupakan fitness functions yang akan dicari nilainya dengan mengoptimumkan atau meminimumkannya dengan mengkuadratkan nilai minimum value atau menggunakan persamaan
5. Kemudian langkahnya selanjutnya mencari Pbest dan Gbest. Pbest merupaakn nilai rata rata dari cluster yang dapat dihitung dengan nilai persamaan (2.2). Gbest adalah nilai keseluruhan Pbest pada semua partikel yang digunakan
6. Setelah ditentukannya nilai Gbest langkah selanjutnya adalah menghitung kecepatan (velocity) dan posisi (position) dengan menggunakan persamaan 2.6 dan persamaan 2.7. dimisalkan pada penelitian ini peneliti menggunakan VO dan $XO = 0$

3. KESIMPULAN

Berdasarkan hasil ekperimen yang peneliti lakukan. Baik pengelompokkan pada pasien stunting dengan menggunakan algoritma klasterisasi k-menas atau nilai akurasi algoritma k-means dioptimasikan dengan pso. Hasil penelitian dengan K-Means dengan menggunakan 3 label menghasilkan label 1 sebagai cluster 1 yang mempunyai nilai 41, dimana label 1 ini merupakan cluster dari stunting berat, label 2 sebagai cluster 2 yang mempunyai nilai 109, dimana label 2 ini merupakan cluster dari stunting dan label 3 sebagai cluster 3 yang mempunyai nilai 50, dimana label 3 ini merupakan cluster dari stunting ringan. Sedangkan hasil penelitian yang dilakukan dengan menggunakan Optimasi K-Means, PSO terbukti bisa meningkatkan nilai akurasi terhadap k-means standar dengan nilai akurasi sebesar 78,88%.

REFERENSI

- Gladys Apriliana, (2018). Analisis faktot – factor resiko terhadap kejadian stunting pada balita 0 – 59 bulan di Negara berkembang dan asia tenggara . universitas Indonesia depok .
- Uliyatul laeli 2019. Pemberdayaan masyarakat terhadap stunting. Universitas nahdhotul ulama Surabaya
- Awa ramdhani (2020) hubungan pengetahuan ibu dengan stunting. Universitas Muhammadiyah tasikmalaya.
- Edwin Daniel osie (2018). “hubungan antara sikap dan pengetahuan ibu terhadap kejadian stunting pada nak baru masuk sekolah dasar di kecamatan nanggolo’
- Han, Jiawet, Kamber Micheline, and Jian Pei. 2012. Data Mining Concepts and Techniques.
- Hartigan, J. a., and M. a. Wong. 1979. “Algorithm AS 136: A K-Means Clustering Algorithm.” *Journal of the Royal Statistical Society C* 28(1): 100–108.
- Bery, A J Micheal, and S Gordon Linoff. 2004. *Data Mining*.
- Bsoul, Qusay, Juhana Salim, and Lailatul Qadri Zakaria. 2013. “An Intelligent Document Clustering Approach to Detect Crime Patterns.” *Procedia Technology* 11(Iceei): 1181–87. <http://dx.doi.org/10.1016/j.protocy.2013.12.311>.