


OPTIMASI PUSAT KLASTER MENGUNAKAN ALGORITMA FAST GENETIC KMEAN PADA DATA BERDISTRIBUSI NORMAL

Budi Nur Iman, Entin Martiana K, Umi
Sa'adah

Politeknik Elektronika Negeri Surabaya (PENS), ITS Surabaya,
Jl. Kampus ITS, Keputih Sukolilo Surabaya, 60111
Email : {alfaruqi, entin, umi}@eepis-its.edu



Abstrak

Penelitian ini membahas kemampuan algoritma fast genetic kmean dalam mengoptimasi pusat klaster. Perbandingan jumlah jarak dalam kelompok (within) dengan jumlah jarak antar kelompok (between) merupakan kriteria klaster. Kriteria klaster menunjukkan hasil optimasi pusat klaster. Semakin kecil kriteria klaster berarti semakin optimal pusat klaster. Hasil penelitian menunjukkan bahwa tingkat kebenaran dari data dengan pola bertumpukan (99,29%), terpisah (99,98%), outlier (98,82%) dan trend (99,2%). Hal ini menunjukkan bahwa optimasi pusat klaster berhasil dilakukan dengan baik oleh algoritma fast genetic kmean.

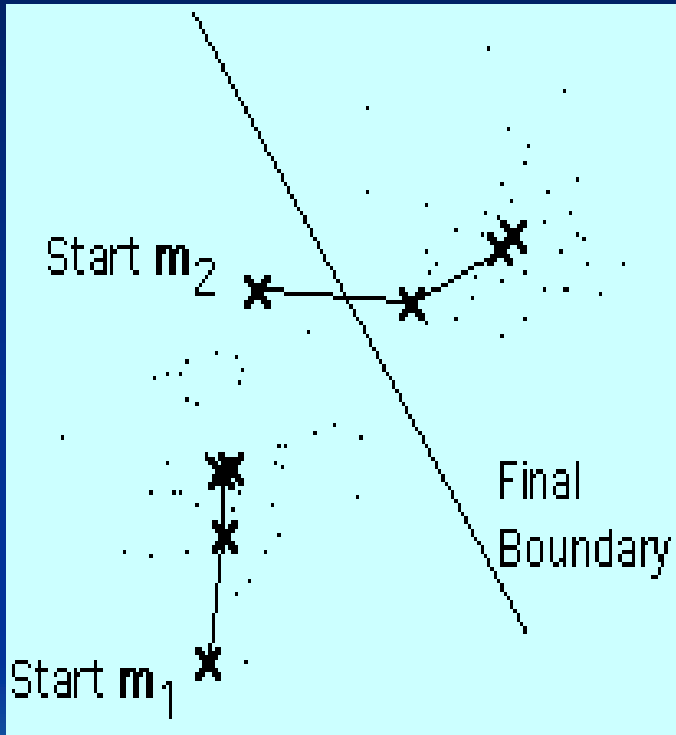
Kata Kunci : fast genetic kmean, variance ratio criterion, tingkat kebenaran

Pendahuluan

- Pengklasteran (*clustering*) merupakan metode untuk mengelompokkan objek, di mana kelompok yang terbentuk diharapkan memiliki karakteristik objek yang serupa dan berbeda karakteristiknya dengan kelompok yang lain.
- Metode kmean menghasilkan klaster yang konvergen tetapi belum tentu optimal.
- Hasil klaster dengan metode kmean dipengaruhi oleh nilai pusat klaster pada saat pembangkitan awal.
- Hasil percobaan yang telah dilakukan, terkadang menghasilkan klaster yang tidak memiliki anggota, sehingga solusi yang diperoleh terjebak pada kondisi lokal minimum.
- Algoritma Fast Genetic Kmean diusulkan untuk diterapkan dalam memperbaiki lokal minimum untuk mendapatkan pusat klaster yang optimal.

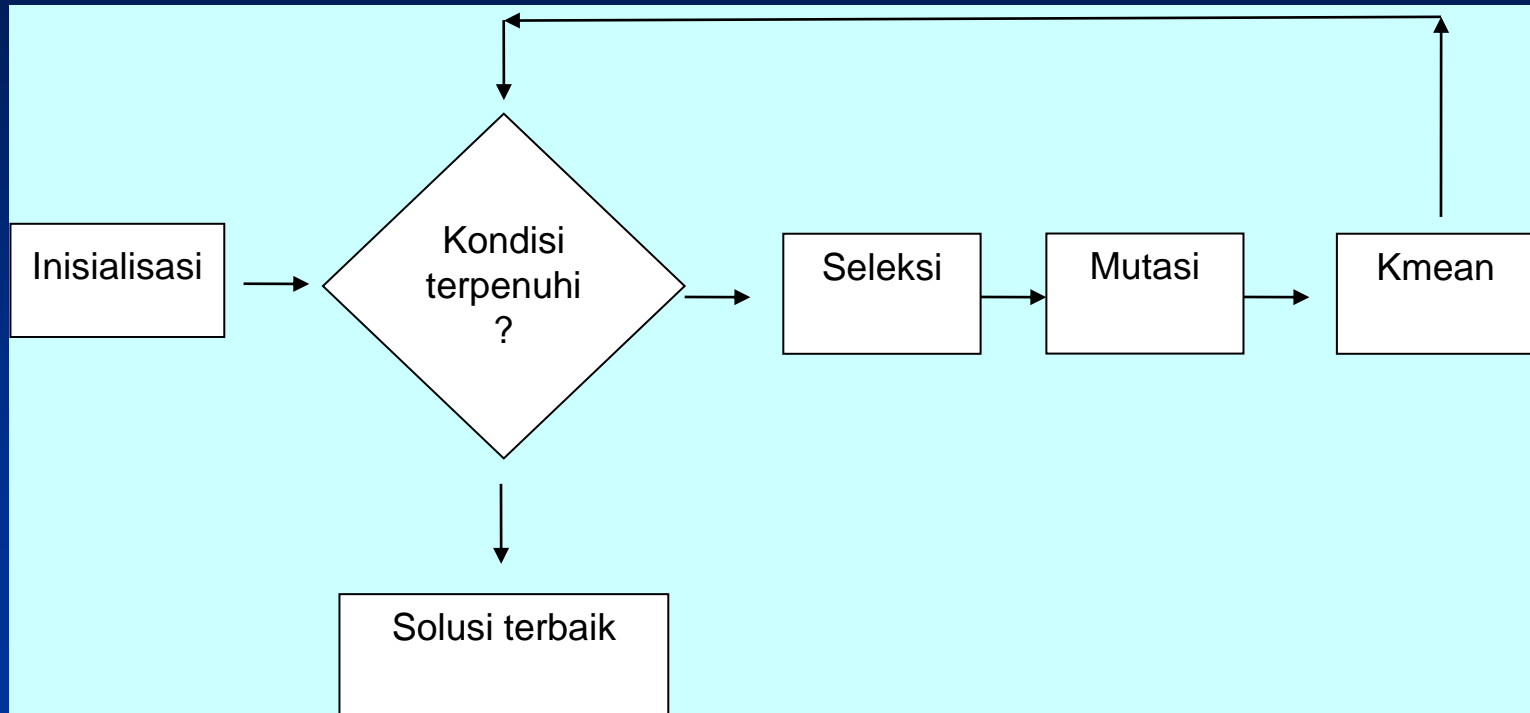


Algoritma Kmean



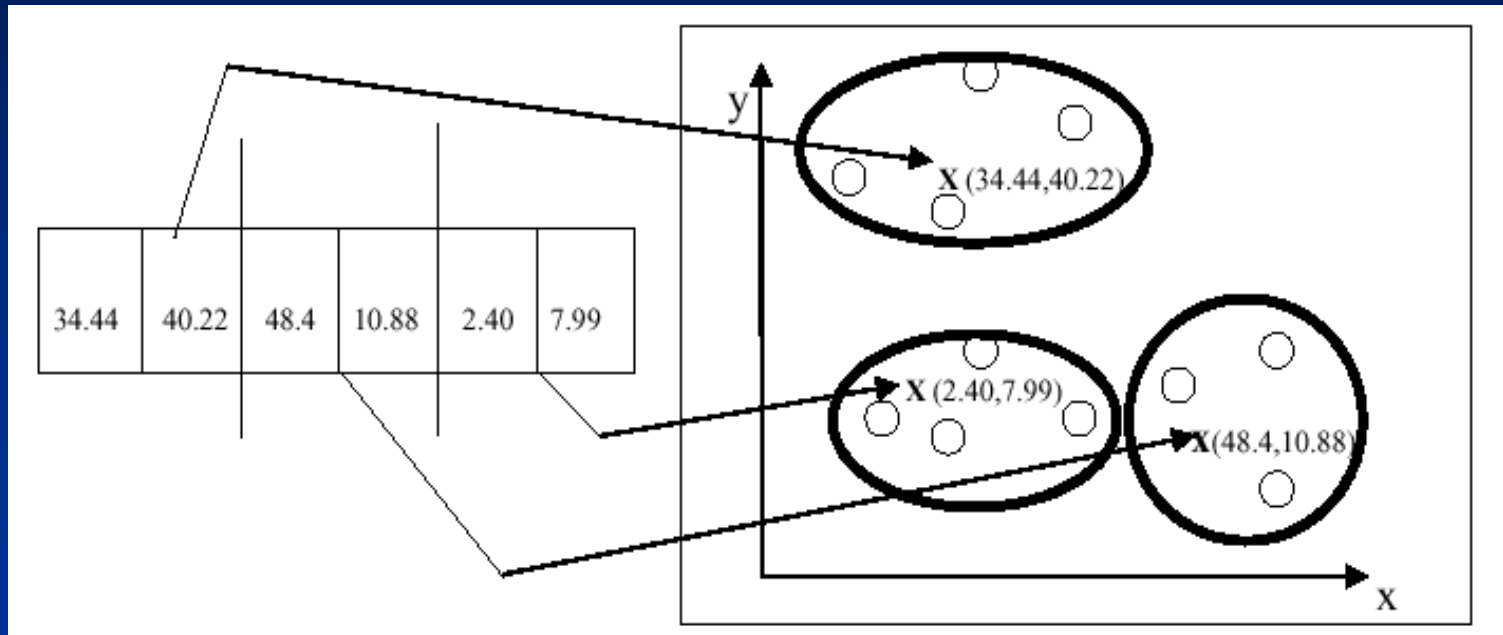
1. Tentukan k buah pusat kelompok awal
2. Hitung jarak tiap data dengan masing-masing pusat kelompok
3. Kelompokkan setiap data ke dalam kelompok terdekat
4. Hitung pusat kelompok sebagai rata-rata data di dalam kelompok
5. Lakukan langkah 2- 4 sampai pusat kelompok tidak berubah

Fast Genetic Kmean



Yi Lu, Shiyong Lu, Farshad Fotouhi, Youping Deng, and Susan Brown, *FGKA : A Fast Genetic K-means Clustering Algorithm*, in Proc. of the 19th ACM Symposium on Applied Computing, pp. 622-623, Nicosia, Cyprus, March, 2004

Representasi Kromosom



Kromosom merupakan kumpulan dari gen yang berupa barisan dari bilangan real yang merepresentasikan K pusat cluster [1][2]. Bila objek yang berisi V variabel, maka panjang kromosom adalah $V \cdot K$

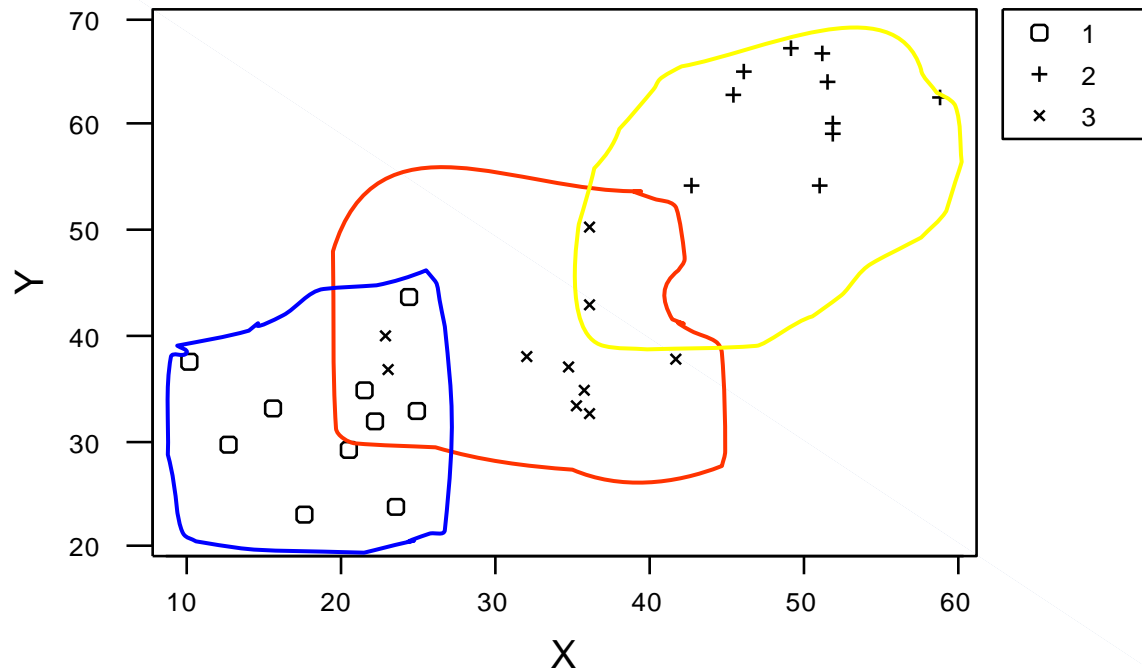
Fungsi Fitnes

$$F = B / W$$

Ukuran yang dapat menyatakan kehomogenan hasil pengklasteran adalah jumlah total jarak dalam kelompok disebut *within* (W) sedangkan ukuran keheterogenan adalah jumlah total jarak antar kelompok disebut juga *between* (B).

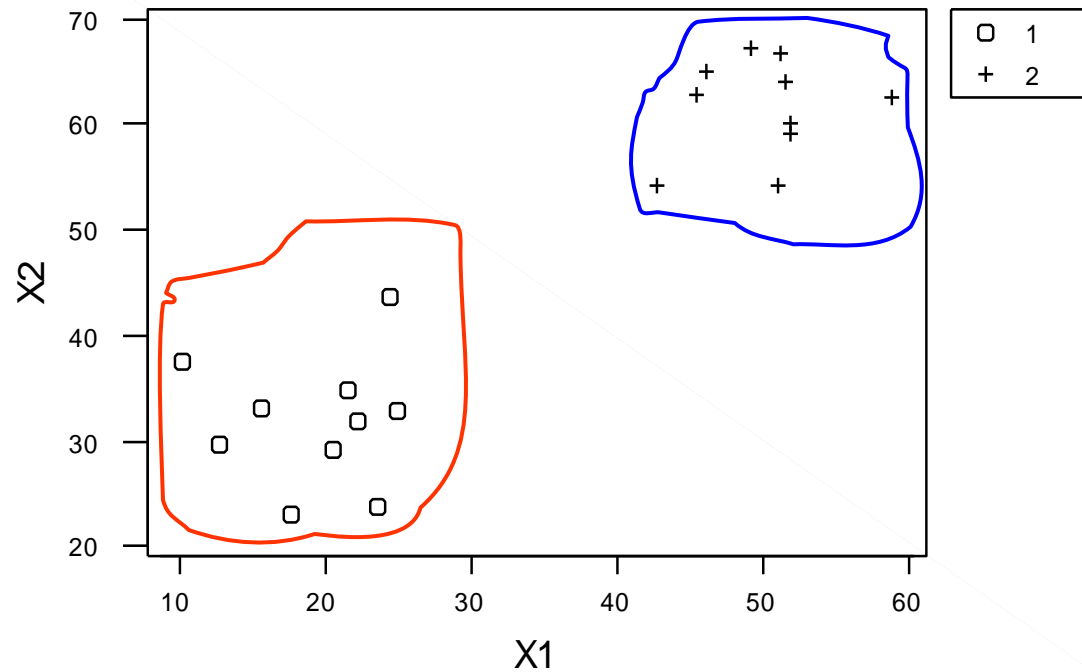
Data bertumpukan

Data Bertumpukan dengan $N = 30$



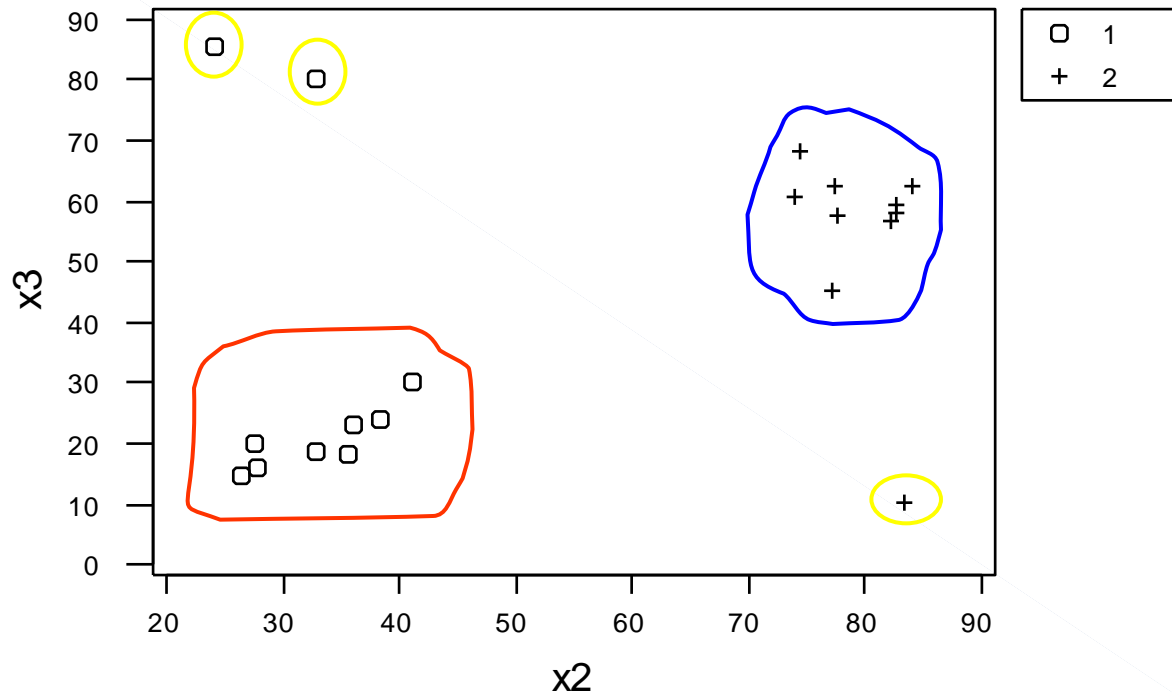
Data Terpisah

Data Terpisah dengan $N = 20$



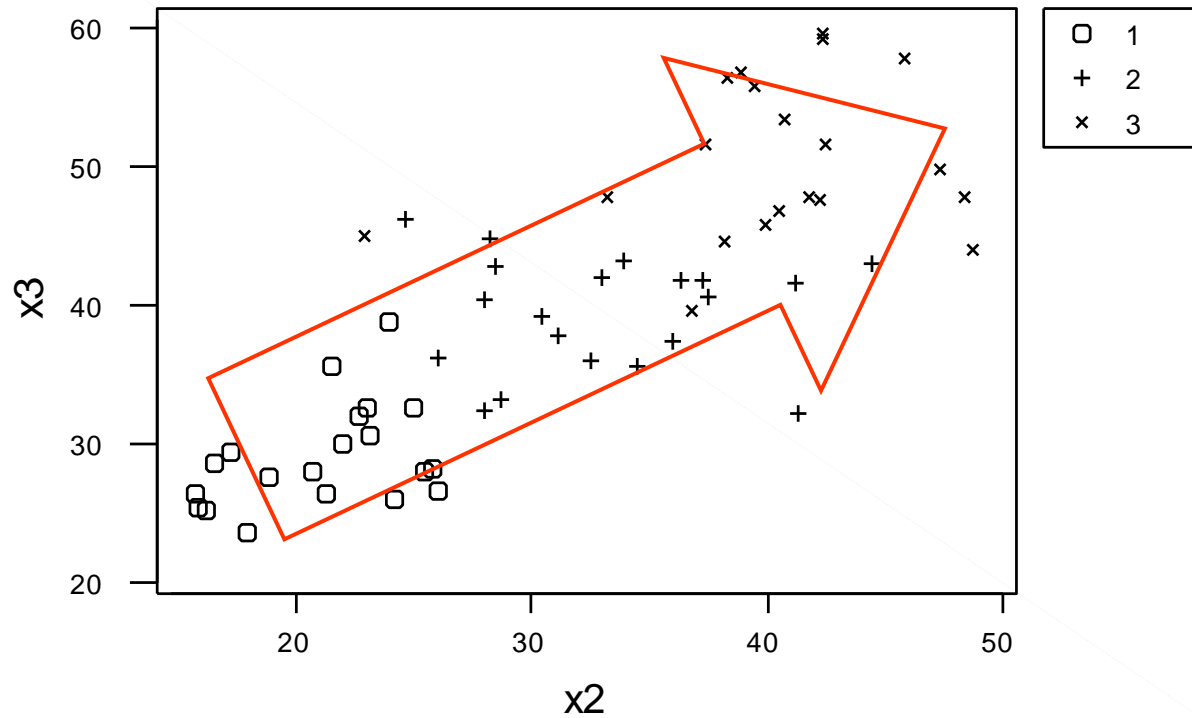
Data Pencilan

Data Pencilan dengan $N = 20$



Data Trend

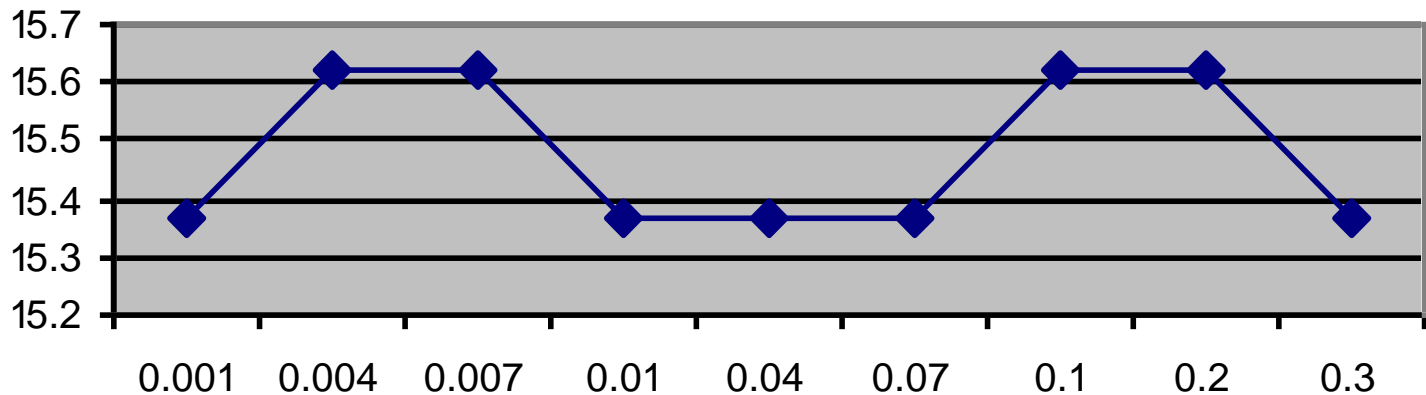
Data Trend dengan N = 60



Perangkat lunak opensource dari University of Tokyo, Institute of Medical Science Human Genome Center (<http://bonsai.ims.u-tokyo.ac.jp>),

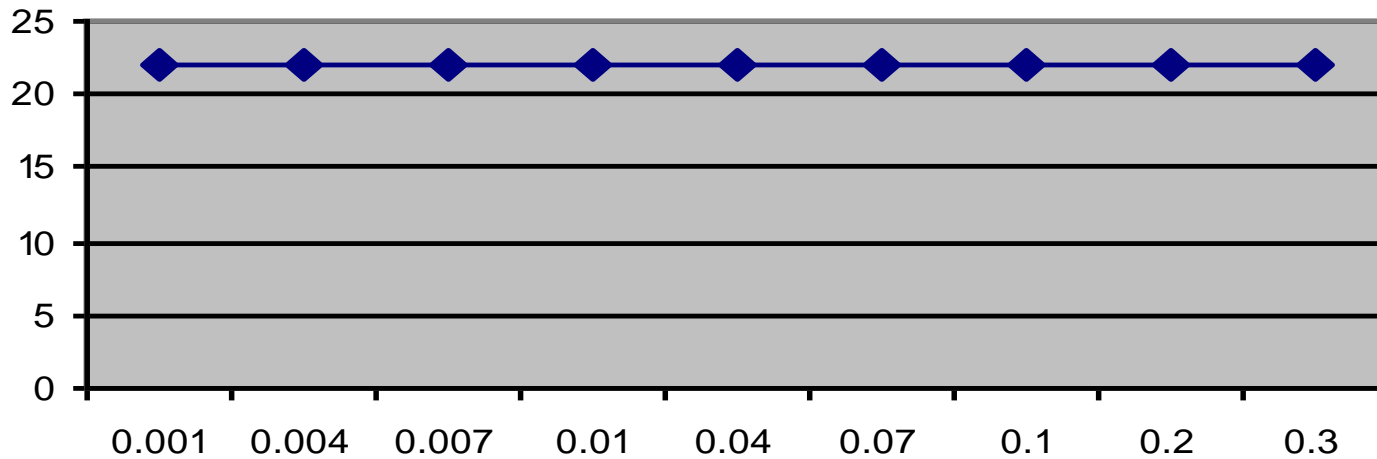
Objek / Data	Within (W)	Between (B)	Perbandingan (W/B)
Bertumpukan	1328.16	86.4	15.37222222
Terpisah	945.62	42.64	22.17682927
Pencilan	8337.04	53.92	154.6186944
Trend	2113.37	60.26	35.07085961

Nilai W/B dari algoritma FGK pada pola Tumpukan



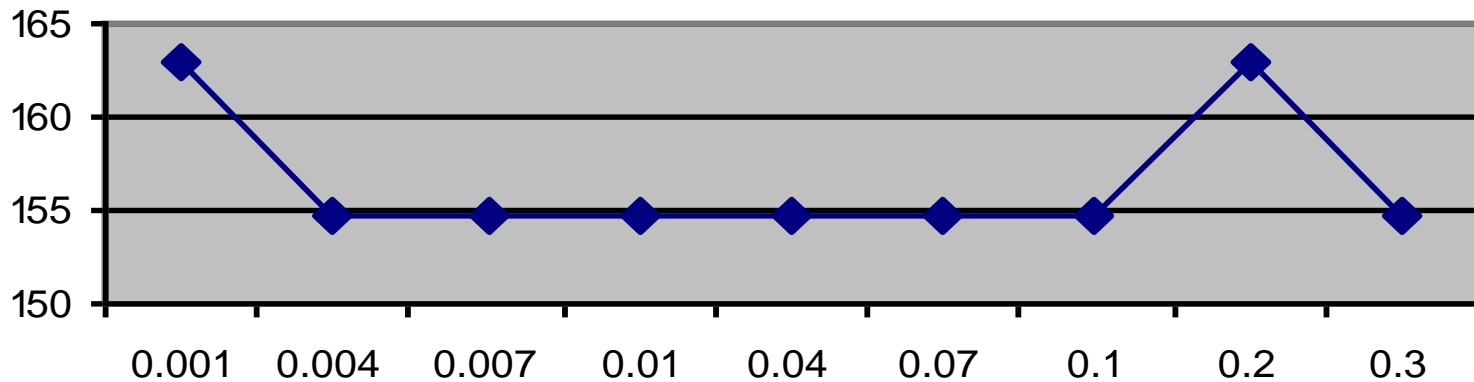
Tampak bahwa pada probabilitas mutasi 0.001, 0.01, 0.04, 0.07 dan 0.3 nilai W/B dengan algoritma FGK lebih kecil dari 15.3722222 dan tingkat kebenarannya 99,29%

Nilai W/B dari algoritma FGK pada pola terpisah



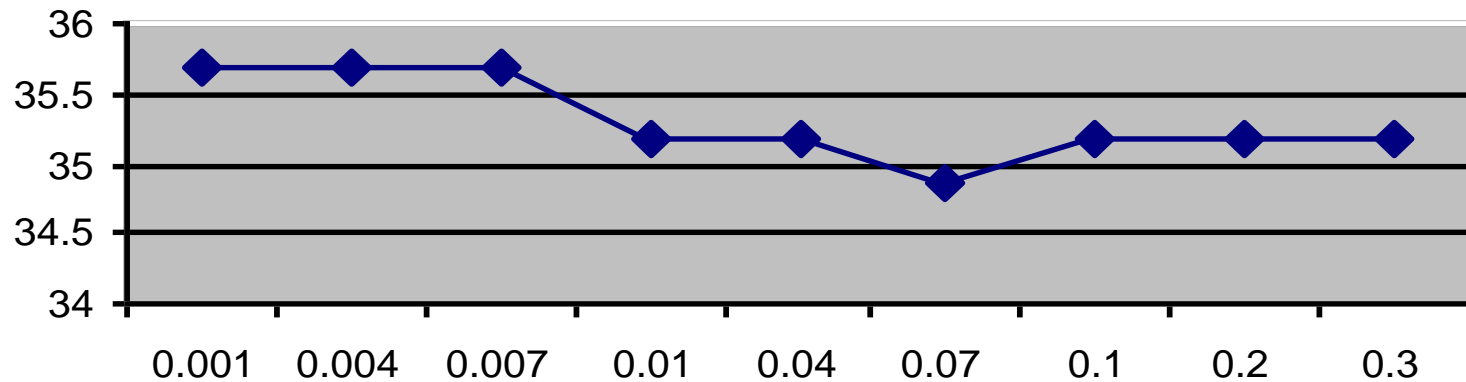
Tampak bahwa pada semua probabilitas mutasi nilai W/B dengan algoritma FGK lebih kecil dari 22.17682927 dan tingkat kebenarannya 99,98%.

Nilai W/B dari algoritma FGK pada pola outlier



Tampak bahwa pada probabilitas mutasi 0.001 dan 0.2 nilai W/B dengan algoritma FGK lebih besar 154.6186944 dan tingkat kebenarannya 98,82%.

Nilai W/B dari algoritma FGK pada pola data trend



Tampak bahwa pada probabilitas mutasi 0.07 nilai W/B algoritma FGK lebih kecil dari 35.07085961 dan tingkat kebenarannya 99,2%.

Kesimpulan

- Tingkat kebenaran nilai W/B dari algoritma fast genetic kmean terhadap metode kmean dari perangkat lunak opensource sangat baik yaitu untuk data bertumpukan (99,29%), terpisah (99,98%), outlier (98,82%) dan trend (99,2%).
- Optimasi pusat klaster berhasil dilakukan dengan baik oleh algoritma fast genetic kmean. Hal ini ditandai dengan diperolehnya nilai W/B yang lebih kecil dari nilai W/B dari perangkat lunak opensource.

Daftar Pustaka

- [1] Anderberg, M.R., *Cluster Analysis for Applications*, New York: Academic Press, 1973
- [2] Andrew Moore: “*K-means and Hierarchical Clustering - Tutorial Slides*” .
Tutorial ada di alamat <http://www-2.cs.cmu.edu/~awm/tutorials/kmeans.html>
- [3] Budi Nur Iman dan Rully Soelaiman, *Penerapan Metode Pengklasteran Hibrida Berbasis Kmeans dan Algoritma Genetika*, Industrial Electronics Seminar (IES), PENS-ITS Surabaya, 2004.
- [4] Budi Nur Iman dan Rully Soelaiman, *Teknik Kluster Menggunakan Algoritma Genetika Pada Data Hasil Ujian Penerimaan Mahasiswa Baru*, ECCIS, Universitas Brawijaya Malang, 2004.
- [5] Cowgill, M., Harvey, R. J., & Watson, *A genetic algorithm approach to cluster analysis*, Computers and Mathematics with Applications, 37, 99-108, 1999
- [6] Chiou, Y. and Lan, L. W. *Genetic clustering algorithms*, European journal of Operational Research, 135:413-427, 2001
- [7] D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Massachusetts, USA, 1989
- [8] Jim Gasvoda dan Qin Ding, *A Genetic Algorithm for Clustering on very large data sets*. Paper ada di http://cs.hbg.psu.edu/~ding/publications/CAINE03_GA.pdf
- [9] Maulik dan S. Bandyopadhyay, *Genetic Algorithm Based Clustering Technique*, Pattern Recognition, vol. 33, no. 9, pp. 1455-1465, 2000.
- [10] Yi Lu, Shiyong Lu, Farshad Fotouhi, Youping Deng, and Susan Brown, *FGKA : A Fast Genetic K-means Clustering Algorithm*, in Proc. of the 19th ACM Symposium on Applied Computing, pp. 622-623, Nicosia, Cyprus, March, 2004