

Comparison of result clustering study case posyandu with the scalable K Means ++ Clustering Method

Ariadi Retno Hayati^{*}, Mamluatul Hani'ah, Ika Kusumaning
Politeknik Negeri Malang

Email Korespondensi : *faniri4education@gmail.com

Abstract. Application in this research to grouping data at unsupervised method, in this study comparing the results grouping with the K mean clustering method, K Means ++ clustering method and the Scalable K Means ++ clustering method. Based on the test results by analyzing the iteration error value, the results of the analysis show that the K Means ++ clustering and Scalable K Means ++ clustering method will produce less error values when compared to the K Means Clustering method. The data used as the basis of analysis in this study is based on data from Posyandu Rajawali Singosari in Malang. The initial initialization value of the centroid can be determined or randomly and is very influential for the data grouping process. Calculation analysis program used scilab programming and the error results with the graph of the minimum value. Result in test data, error value test data 1 get Scalable K Means ++ clustering error minimum 0,07, test data 2 get error value minimum K Means ++ Clustering 0,15, test data 3 get error value minimum 0,005 at metode Scalable K Means Clustering, test data 4 get error value minimum 0,15 at K Means ++ Clustering.

Keywords : Scalability K Means ++, K Means ++ Clustering, K Means Clustering, Data Posyandu Rajawali Singosari

1. Pendahuluan

Metode pembelajaran dengan pengelompokan (*clustering*) data pada penelitian ini diterapkan pada data posyandu Rajawali kecamatan Karangploso Kabupaten Malang, dimana data dari data Posyandu akan dikelompokkan oleh metode *clustering*. Metode *clustering* pada penelitian ini akan membandingkan metode Scalable K Means ++ untuk menentukan kecenderungan pengelompokkan berdasarkan pemeriksaan bulanan dari data posyandu. Metode *clustering* adalah bagian dari *Machine Learning*. Permasalahan pada pembelajaran *unsupervised* yaitu mengidentifikasi data dan pengelompokkan data. *Machine Learning* [3] adalah pembelajaran sistem secara otomatis.

Proses pembelajaran dimulai dari observasi data pada penelitian ini yaitu data posyandu terdapat data berat, data IMT. Penentuan hasil pengelompokan oleh *unsupervised machine learning* [9] ditentukan berdasarkan pembelajaran pada setiap iterasinya dimana pada metode *clustering* akan memperbarui nilai pembelajaran pada setiap iterasinya. Pada penelitian sebelumnya, oleh Windha Mega Pradnya Duhita

dengan judul Clustering Menggunakan Metode K Means Untuk Menentukan Status Gizi Balita dari STMIK Amikom Yogyakarta [2], mengelompokkan data balita dari Puskesmas Karang Songo dengan menerapkan metode K Means Clustering data. Penerapan metode *K Means clustering* pada data obat-obatan oleh Gustientiedina, Hasmil Adiya, Yenny Desnelita, dengan judul Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan[4] pada RSUD Pekanbaru dari Sekolah Tinggi Ilmu Komputer Pelita Indonesia, menerapkan metode *K Means clustering* untuk pengelompokkan jumlah kebutuhan obat rumah sakit. Penerapan metode *K Means Clustering* untuk pengelompokkan data penjualan [5] oleh Benri Melpa Metisen, Herlina Latipa Sari dengan judul Analisis Clustering Menggunakan Metode K Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila, dari Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Dehasen, yaitu mengelompokkan statistika data hasil penjualan dari swalayan.

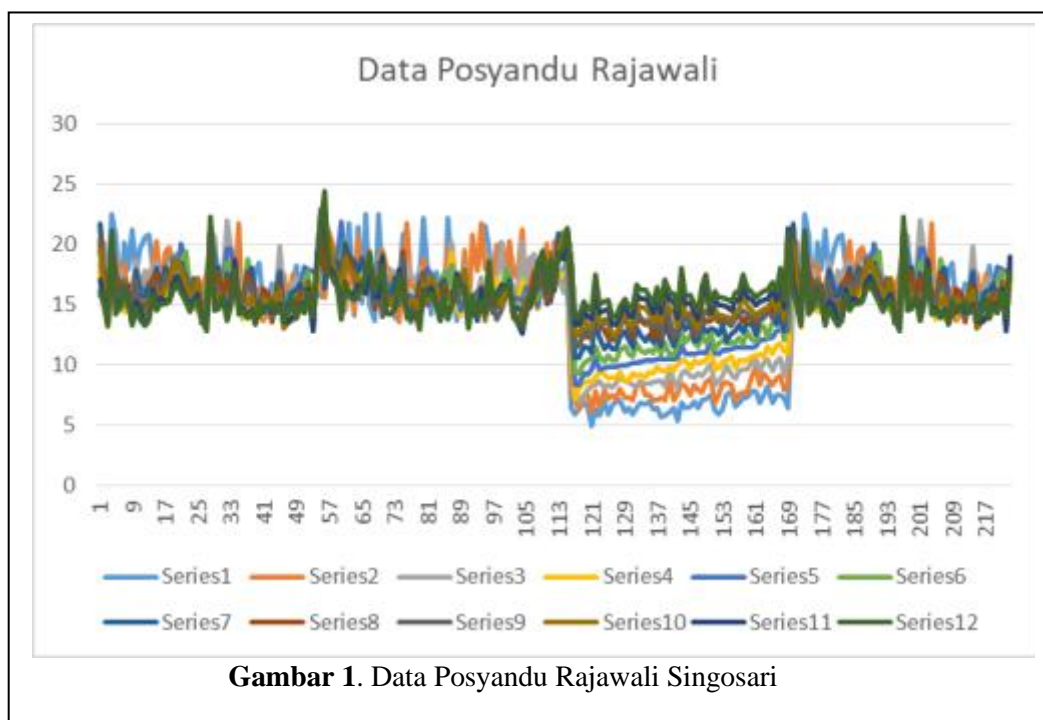
Pada penelitian ini menerapkan pada data Posyandu dengan harapan hasil inisiliasi dari *cluster* memiliki karakteristik dengan menghasilkan nilai kesalahan yang lebih rendah dengan menggunakan metode Scalable K Means ++ Clustering [1].

2. Metode Penelitian

Pada penelitian ini menggunakan metode *Scalable K Means ++* yang merupakan pengembangan metode *K Means Clustering* dan *K Means ++ Clustering*. Pada pengelompokkan data terdapat penghitungan untuk menentukan kriteria data pengelompokkan, yaitu dengan rumus *eucledian* [8][10]. Untuk menentukan nilai titik pusat dan jumlah pengelompokkan akan dipebarui pada setiap iterasinya[6].

2.1 Pengumpulan Data

Data untuk data mining dapat diterapkan pada berbagai bidang dan tujuan [11], data yang digunakan pada penelitian ini adalah data Posyandu yaitu data berat tubuh sebagaimana grafik pada Gambar 1.

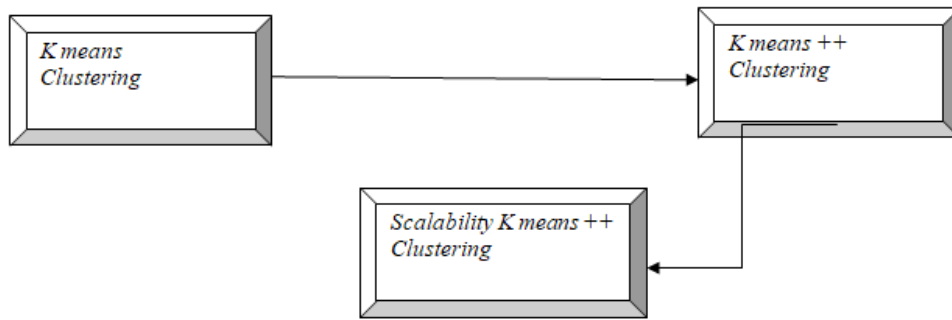


Data berat badan balita didata oleh kader posyandu, secara kontinyu setiap bulannya, dimana data tersebut akan dikelompokkan menjadi data kesehatan normal, diatas normal dan dibawah nilai normal.

2.2 K Means Clustering

Pada penelitian ini menerapkan data posyandu untuk meneliti nilai kesalahan yang didapatkan dengan

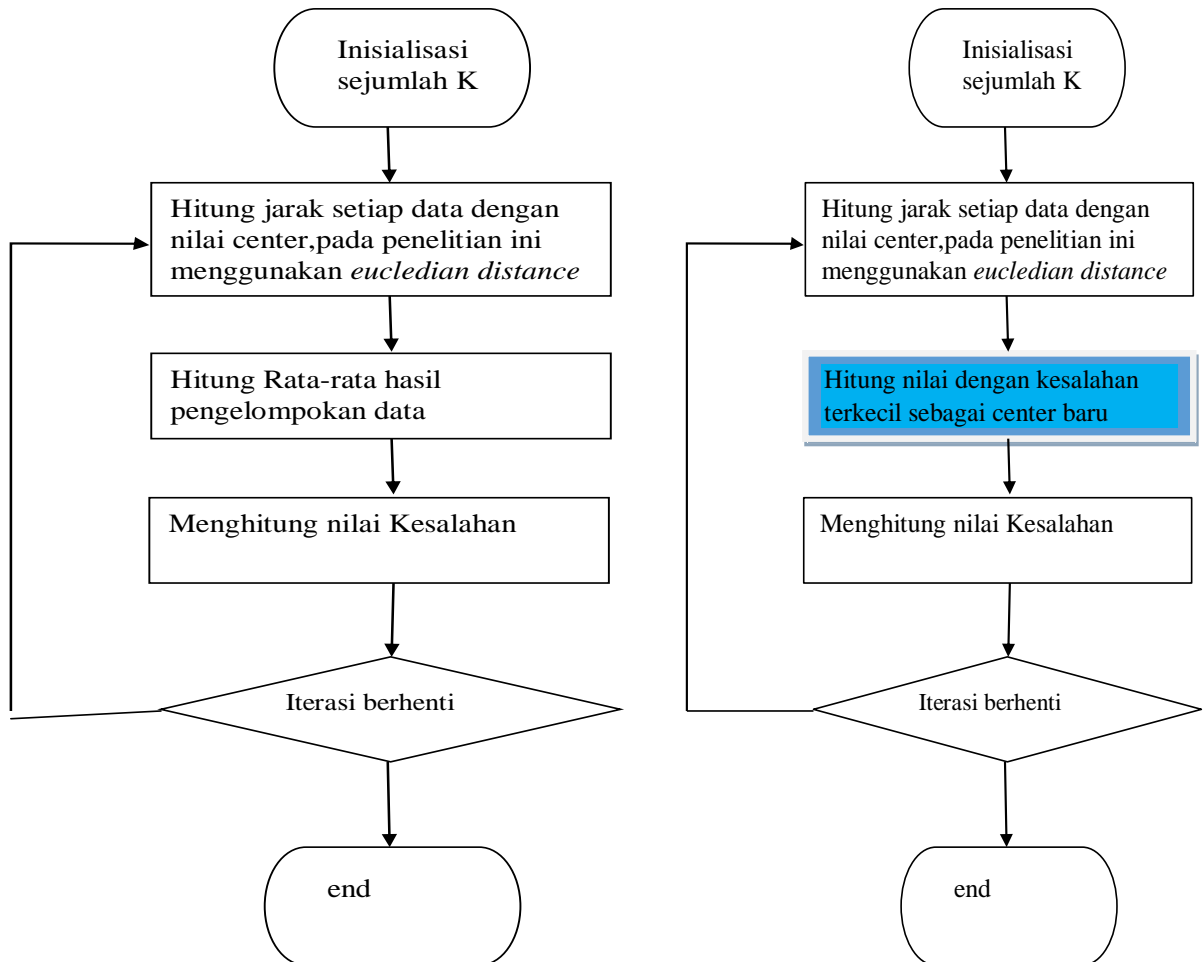
metode clustering.



Gambar 2. Pengembangan Metode *Clustering* Berdasarkan Metode *K Means Clustering*

2.2.1 *K Means ++ Clustering*

Metode *K Means Clustering ++* adalah metode pengelompokan data dengan menerapkan inisial jumlah kelompok sejumlah *K* kelompok dan di kelompokkan pada setiap iterasinya dengan berdasarkan jarak terdekat [12] dapat dilihat pada gambar 3.



a. Metode *K Means Clustering Clustering*

b. Metode *K Means ++ Clustering*

Gambar 3. Diagram Flowchart Metode *K Means Clustering* dan *K Means ++ Clustering*

Menghitung jumlah kesalahan dengan menerapkan rumus untuk menganalisa kesalahan setiap data yang terkelompok pada *cluster* ke- i dimana i dari jumlah K *cluster*. *Error* data-data terhadap masing - masing center (J_k) sebagai yaitu :

$$J_k = \sum_{x \in P_k} \|x_j - m_k\|^2 \quad (1)$$

Keterangan:

J_k = nilai kesalahan pada kelompok *cluster* ke - k

x = data Posyandu

P_k =Data-data anggota pada kelompok cluster ke- k

m_k =nilai *center* pada *cluster* ke- k .

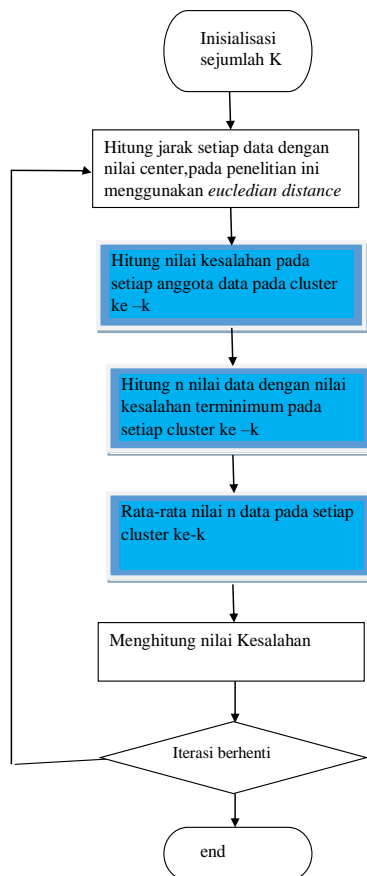
Pada setiap iterasi akan dihitung jumlah nilai J_k , dan akan dibandingkan hasilnya dengan pengembangan metode *K Means Clustering*, pada sub bab 2.2.2 akan membahas mengenai *K Means ++ Clustering* . Pada gambar 3.b, yang berbeda pada penghitungan *K Means ++ Clustering* dibandingkan *K Means Clustering* adalah tahap pembaharuan nilai *cluster* pada setiap iterasinya. Pembaharuan nilai *cluster* pada setiap iterasinya dengan mencari nilai kesalahan terminimum dengan rumus J_k sehingga nilai *center* yang baru pada iterasi ke- i adalah nilai dari data dengan nilai kesalahan terminimum. Penghitungan nilai kesalahan pada setiap *cluster* ke k adalah sebagaimana berikut:

$$d = \|x_j - m_k\|^2 \quad (2)$$

Keterangan:

d =jarak pada setiap data pada *cluster* ke- k

x = nilai data



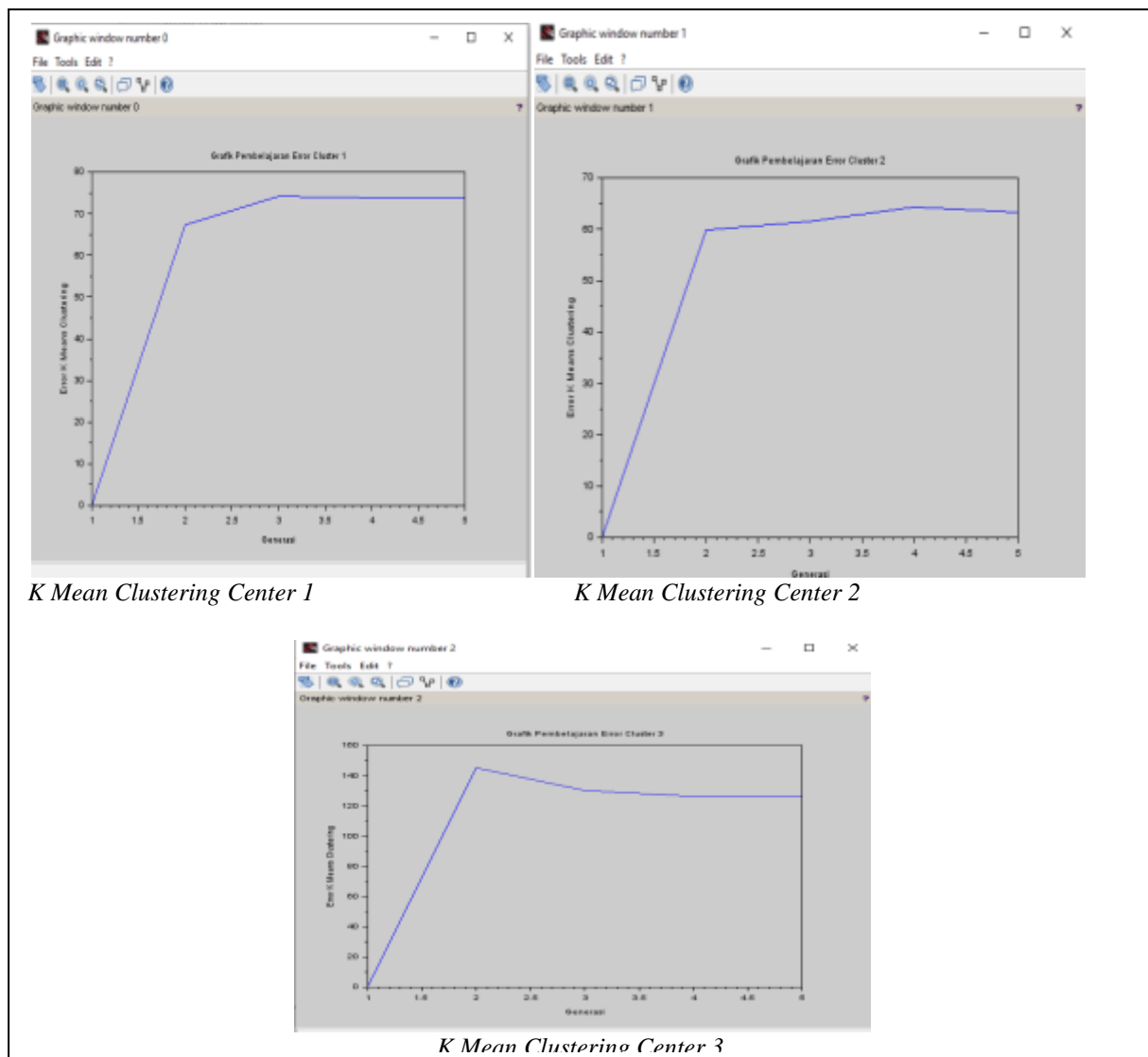
Gambar 4. Diagram Flowchart Metode *Scalable K Means++ Clustering*

Scalable K Means++ Clustering ada metode *Scalable K Means++ clustering* memiliki konsep yang sama dengan metode *K Means Clustering* dan *K Means ++ Clustering* karena merupakan pengembangan metode dari metode *K Means ++ Clustering*.

Pada metode *Scalable K Means ++ Clustering*, terdapat perbedaan dengan metode *K Means ++ Clustering* yaitu terdapat penentuan jumlah n sebagai data yang akan diacu dengan nilai kesalahan minimum dan akan dirata-rata nilai kesalahannya pada setiap *center*.

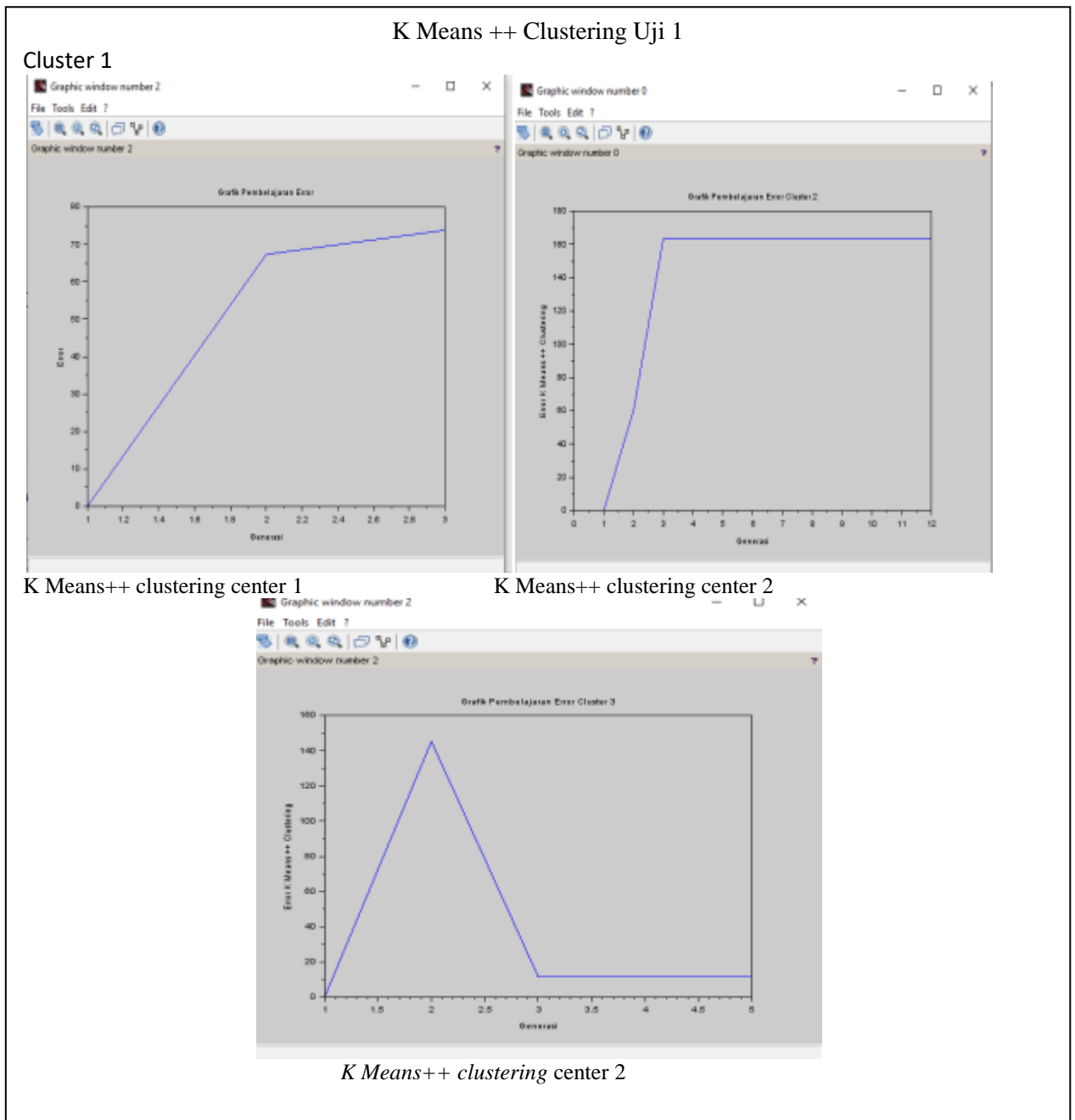
3. Hasil Dan Pembahasan

Pada penelitian ini akan membahas hasil penghitungan setiap *cluster* atau kelompok dengan jumlah dimensi pemeriksaan bulanan, kesalahan pengelompokan data dengan menerapkan metode *K Means*



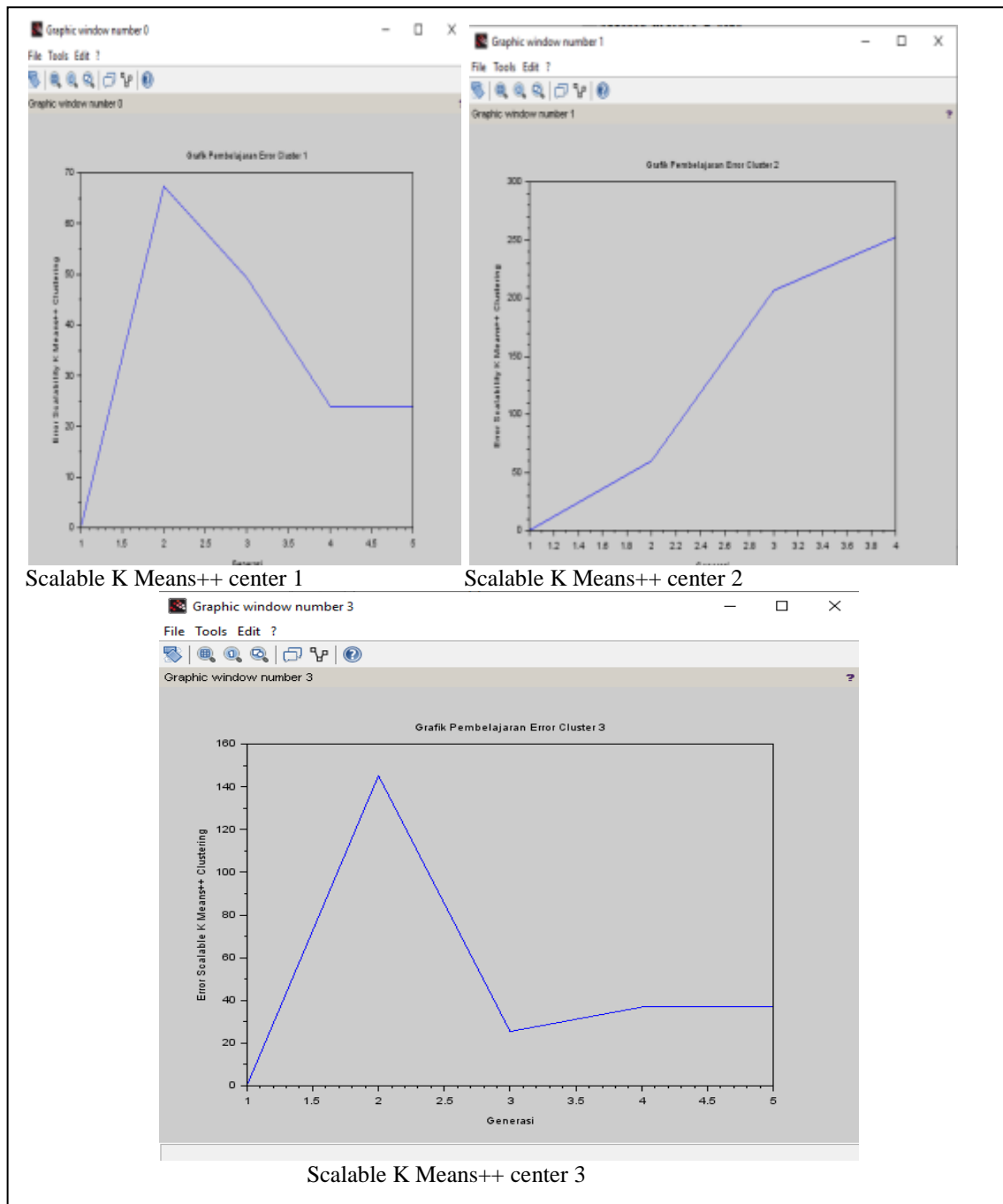
Gambar 5. *K Means Clustering Uji 1*

Clustering, K Means ++ Clustering dan Scalable K Means ++ Clustering. Pada gambar 5, merupakan data uji pada metode *K Means Clustering* dimana jumlah kesalahan pada *cluster 1* adalah 0,28 dari jumlah kesalahan *cluster*, *cluster 2* adalah 0,24 dari kesalahan keseluruhan, dan *cluster 3* adalah 0,47 dari jumlah kesalahan *cluster*.



Gambar 6. *K Means ++ Clustering Uji 1*

Pada gambar 6, jumlah kesalahan pada *cluster 1*, adalah 0,29 dari jumlah kesalahan keseluruhan, jumlah kesalahan pada *cluster* adalah 0,65 dari jumlah kesalahan keseluruhan, dan pada *cluster* memiliki nilai kesalahan 0,04 dari jumlah kesalahan keseluruhan pada *cluster*.



Gambar 7. Scalable K Means ++ Clustering Uji 1

Pada gambar 7, jumlah kesalahan pada *cluster 1*, adalah 0,07 dari jumlah kesalahan keseluruhan, jumlah kesalahan pada *cluster 2* adalah 0,805 dari jumlah kesalahan keseluruhan, dan pada *cluster 3* memiliki nilai kesalahan 0,11 dari jumlah kesalahan keseluruhan pada *cluster* . Hasil penelitian sebagaimana sesuai dengan tabel 1,2,3 dan 4 yaitu tabel uji coba hasil pengelompokan data.

Tabel 1. Uji Coba Data 1

| Uji Coba 1 | Cluster 1 | Cluster 2 | Cluster 3 |
|------------------------------|-----------|-----------|-----------|
| K Mean Clustering | 0,28 | 0,24 | 0,47 |
| K Mean ++ Clustering | 0,29 | 0,65 | 0,04 |
| Scalable K Mean++ Clustering | 0,07 | 0,805 | 0,11 |

Tabel 2. Uji Coba Data 2

| Uji Coba 2 | Kesalahan cluster ke 1 | Kesalahan cluster ke 2 | Kesalahan cluster ke 3 |
|-------------------------------|------------------------|------------------------|------------------------|
| Metode K Means Clustering | 0.3504131 | 0.2040088 | 0.4455781 |
| K Means ++ Clustering | 0.54553 | 0.1578865 | 0.2965835 |
| Scalable K Means++ Clustering | 0.3504131 | 0.2040088 | 0.4455781 |

Tabel 3. Uji Coba Data 3

| Uji Coba 3 | Kesalahan cluster ke 1 | Kesalahan cluster ke 2 | Kesalahan cluster ke 3 |
|-------------------------------|------------------------|------------------------|------------------------|
| Metode K Means Clustering | 0.1082261 | 0.5376094 | 0.3541645 |
| K Means ++ Clustering | 0.0475596 | 0.4533808 | 0.3101942 |
| Scalable K Means++ Clustering | 0.0572949 | 0.696537 | 0.2461682 |

Tabel 4. Uji Coba Data 4

| Uji Coba 4 | Kesalahan cluster ke 1 | Kesalahan cluster ke 2 | Kesalahan cluster ke 3 |
|-------------------------------|------------------------|------------------------|------------------------|
| Metode K Means Clustering | 0.0572949 | 0.696537 | 0.2461682 |
| K Means ++ Clustering | 0.54553 | 0.1578865 | 0.2965835 |
| Scalable K Means++ Clustering | 0.3504131 | 0.2040088 | 0.4455781 |

4. Kesimpulan

Hasil uji coba berdasarkan tabel pada tabel 1, 2 3 dan 4 maka metode *K Means ++ Clustering* dan metode *Scalable K Means ++ Clustering* memiliki kecenderungan memiliki jumlah kesalahan center yang lebih minimum dibandingkan metode *K Means Clustering*.

5. Ucapan Terima Kasih

Pada penelitian ini kami ucapkan terimakasih pada Politeknik Negeri Malang dan instansi Posyandu Rajawali Singosari atas mendukung data pada penelitian ini.

6. Daftar Pustaka

- [1] Bahmani,B.,Moseley, B., Vattani, A., Kumar, R., Vassilvitskii, S. (2012).. Scalable K Means++, PVLDB, Vol. 5, No. 7, pp. 622-633
- [2] Dhuhita,W, M, P.(2015) , Clustering Menggunakan Metode K Means Untuk Menentukan Status Gizi Balita, Jurnal Informatika, Vol. 15, No. 2.
- [3] Roihan, A., Sunarya, A., Rafika, A, S. (2019), Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper, Indonesian Journal on Computer and Information Technology, 5.
- [4] Gustientiedina., Adiya, M, H, Desnelita, Y., (2019), Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru, Jurnal Nasional Teknologi Dan Sistem Informasi- Vol. 05 No. 01.
- [5] Metisen, B, M.,Sari, H, L .,(2015), Analisis Clustering Menggunakan Metode K Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila, Jurnal Media Infotama Vol. 11 No 2.
- [6] Lida, M.,(2018) Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. . DKI Jakarta Dengan K-MEANS.
- [7] Wakhidah, N.,(2010) , Clustering Menggunakan K-M algorithm , Jurnal Transformatika Vol 8 No 1.
- [8] Juliawan, D., Amir, F., Desi, E., (2019), Penerapan Data Mining Metode Clustering Pada CV. Secom Infotech Menggunakan Algoritma KMeans, Prosiding Seminar Nasional Riset Information Science, ISSN: 2686-0260.
- [9] Nurhayati., Busman., Iswara, R, P., (2019), Pengembangan Algoritma Unsupervised Learning Technique Pada Big Data Analysis di Media Sosial Sebagai Media Promosi On Line Bagi Masyarakat, Jurnal Teknik Informatika Vol 12 No. 1
- [10] Anggara, M., Sujiani, H., Nasultion, H., (2016), Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokkan Member Di Alvaro Fitness, Jurnal Sistem dan Teknologi Informasi (JUSTIN) Vol. 1, No. 1.
- [11] Thamrin, S.A., (2006), Penggunaan Data Mining Saat Ini Dan Tantangannya Di Masa Depan, Jurnal atematika Statistika dan Komputasi, Vol 3 No 1.
- [12] Fikri, M,C., Agustin, M, E,F., Mintarsih, F.(2017), Pengelompokkan Kualitas Kerja Pegawai Menggunakan Algoritma K-Means++ Dan Cop-KMeans Untuk Merencanakan Program Pemeliharaan Kesehatan Pegawai di PT. PLN P2B JB Depok, Jurnal Psudocode Vol 4 No 1.

