

ANALISIS ALGORITMA K-MEDOIDS CLUSTERING DALAM PENGELOMPOKAN PENYEBARAN COVID-19 DI INDONESIA

Sukma Sindi, Weni Ratnasari Orktapia Ningse, Irma Agustika Sihombing, P.P.P.A.N.W
Fikrul Ilmi R.H.Zer, Dedy Hartama

*Program Studi Teknik Informatika, STIKOM Tunas Bangsa Pematangsiantar,
Jl. Jend. Sudirman Blok A No.1,2 dan 3, Kota Pematangsiantar 21111 - Indonesia
skmsindi@gmail.com, weni.oktapia21@gmail.com, Irmasihombibg006@gmail.com,
fikrulilmizer@gmail.com, dedyhartama@amiktunasbangsa.ac.id*

Abstract - At the beginning of March Indonesia was entering the corona outbreak virus (COVID) Every day the case of Covid-19 distribution in Indonesia continued to increase. the community is issued to conduct social distance to cut the distribution of COVID-19 distribution distributed in various regions. In Indonesia, therefore, the data that has been accommodated is certainly a lot, from the data it can be seen patterns - selection patterns of distribution of COVID-19 distribution are based on test scores, This study uses the *K-Medoids* method so that the distribution patterns of COVID-19 distribution can be used for the community. *K-Medoids* is a method of grouping Analytical sections that aim to get a set of *k*-clusters among the data that most require an object in the collection of data. The results of the new COVID-19 research grouping show the community produced from various regions in Indonesia. Characteristics with a body temperature above 36.9 ° c and with fever and cough resolution supported by one of the characteristics of COVID-19 symptoms.

Kata Kunci - *K-Medoids* Algorithm, Clustering, Data Mining, COVID-19, Data Grouping

Abstrak - Pada awal maret Indonesia sedang di landa masuknya wabah *virus corona (covid)* Setiap hari kasus penyebaran *covid-19* di indonesia terus meningkat. masyarakat diminta untuk melakukan *social distancing* guna mamutus rantai penyebaran *covid-19* yang tersebar diberbagai wilayah.di Indonesia. Oleh karena itu, data yang telah ditampung pastinya banyak sekali, dari data tersebut dapat dilihat pola – pola penentuan pengelompokan penyebaran *covid-19* dilakukan berdasarkan nilai tes, Penelitian ini menggunakan metode *K-Medoids* agar dapat diketahui pola pemilihan penentuan pengelompokan penyebaran *covid-19* bagi masyarakat. *K-Medoids* merupakan metode Analitis *partisional clustering* yang bertujuan untuk mendapatkan suatu set *k-cluster* di antara data yang paling mendekati suatu objek dalam pengelompokan suatu data.. Hasil penelitian pengelompokan penyebaran *covid-19* baru menunjukkan bahwa masyarakat yang berasal dari berbagai wilayah di Indonesia. Ciri-ciri dengan suhu badan di atas 36,9° c dan dengan disertai demam dan batuk berkelanjutan menunjukkan salah satu ciri-ciri gejala *covid-19*

Kata Kunci - Algoritma *K-Medoids*, Clustering, Data Mining, Covid-19, Pengelompokan Data

I. PENDAHULUAN

Diawal maret 2020 Indonesia sedang dilanda pandemi *covid-19*, virus ini pertama kali muncul di wilayah wuhan-tiongkok ditemukan pada akhir desember 2019. pemerintah dan masyarakat sedang melakukan *social distancing* guna memutus rantai penyebaran *covid-19* yang terus meningkat setiap harinya. Penularan *Covid-19* hanya bisa dicegah dengan kedisiplinan untuk tetap berada di rumah. Risiko penularan pun bisa terhindarkan dan masalah Covid-19 di Indonesia bisa segera diatasi kebijakan yang paling baik. Total keseluruhan penyebaran *covid-19* di Indonesia saat ini mencapai 14,265 jiwa untuk kasus penyebaran sembuh *covid-19* di Indonesia mencapai 2.881 jiwa dan angka kematian mencapai 991 jiwa. Penelitian ini menggunakan metode

Penerapan *K-Medoids* agar dapat diketahui pola pemilihan penentuan pengelompokan penyebaran *covid-19* di berbagai wilayah di indonesia. *K-Medoids* merupakan metode *partisional clustering* dimana bertujuan untuk menemukan satu set *k-cluster* di antara data yang paling mencirikan objek dalam kumpulan suatu data.

A. Data Mining

Data mining adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besars. *Data mining* adalah analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut[1].

Descriptive mining, yaitu proses untuk menemukan karakteristik penting dari data dalam satu basis data. Teknik data mining yang termasuk *descriptive mining* adalah *clustering*, *asosiation*, dan *sequential mining* [2].

B. Clustering

Clustering merupakan suatu proses pengelompokkan record suatu , observasi, atau mengelompokkan kelas yang memiliki kesamaan objek. Perbedaan *clustering* dengan *klasifikasi* yaitu tidak adanya variabel target dalam melakukan suatu pengelompokan pada proses *clustering*. *Clustering* sering dilakukan sebagai untuk langkah awal dalam proses data mining saat melakukan suatu metode analisis. Terdapat banyak algoritma *Clustering* yang telah digunakan oleh peneliti sebelumnya seperti *K-Means*, *Improved K-Means*, *Fuzzy C-Means*, *DBSCAN*, *K-Medoids (PAM)*, *CLARANS* dan *Fuzzy Subtractive*. Setiap algoritma memiliki kelebihan dan kekurangan masing-masing, namun prinsip algoritma sama, yaitu mengelompokkan data sesuai dengan karakteristik dan mengukur jarak kemiripan antar data dalam satu kelompok [3].

C. Algoritma K-Medoids

K-Medoids merupakan suatu algoritma yang digunakan untuk menemukan *medoids* didalam sebuah kelompok (*cluster*) yang merupakan titik pusat dari suatu kelompok (*cluster*). Algoritma *K-Medoids* lebih baik dibandingkan dengan *K-Means* karena pada *K-Medoids* kita menemukan *k* sebagai objek yang representatif untuk meminimalkan jumlah ketidaksamaan objek data, sedangkan pada *K-Means* menggunakan jumlah jarak *euclidean distances* untuk objek data[4]. Langkah-langkah algoritma *K-Medoids* sebagai berikut:

1. Inisialisasi pusat *cluster* sebanyak *k* (jumlah *cluster*).
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan ukuran jarak *Euclidean Distance* dengan persamaan: x

$$\sum_{i=1}^n (x_i - y_i)^2; 1, 2, 3, \dots . nd$$

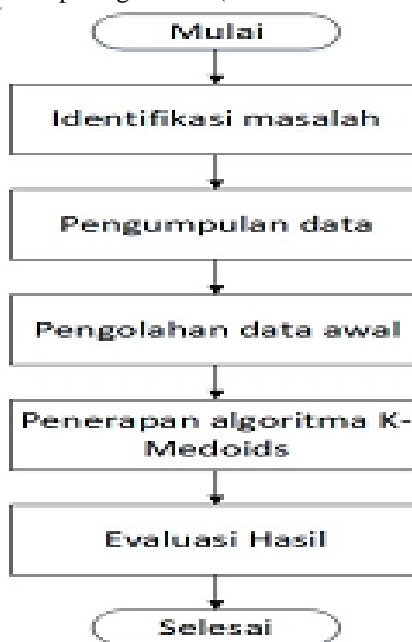
$$(x , y) = \sqrt{x - y} = \sqrt{x} \dots\dots\dots(1)$$

3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoid* baru.
4. Hitung jarak setiap objek yang berada pada setiap masing-masing *cluster* dengan menempuh *medoids* baru.
5. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru – total *distance* lama.

6. Jika $S < 0$, maka ganti objek dengan data *cluster* untuk memperoleh sekelompok *k* objek yang baru sebagai *medoids*.
6. Ulangi tahap ke 3 sampai 5 hingga tidak terjadi perubahan *medoid*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing. Kemudian untuk mendapatkan nilai k di sebuah data yang ada di *clustering K-Medoid* dilakukan di dalam proses *clustering* dapat dipilih dengan berdasarkan nilai DBI (*Davies Bouldin Index*) terkecil.

II. METODE PENELITIAN

Dalam metode ini cara untuk memperoleh dan mengumpulkan data-data ilmiah yang dibutuhkan dengan fungsi dan tujuan tertentu. Metode penelitian dapat dilihat pada gambar 1 (Juninda and Andri, 2019)



Gambar 1. Metode Penelitian

A. Identifikasi Masalah

Dalam penelitian ini, identifikasi masalah bertujuan untuk mengetahui seberapa banyak pengelompokan pada masalah kasus covid-1 di Indonesia menggunakan metode *K-Medoids Clustering* . Dan mencari akar masal dengan table data yang kita punya dengan table yang kita miliki tersebut melonjaknya angka penyebaran wabah menjadi akar pokok permasalahan ini.

B. Pengumpulan Data

Pada pengumpulan Data dilakukan dari objek yang kita telitih dimana data yang kita peroleh dari Kementerian Kesehatan Indonesia. Data yang

digunakan menggunakan 3 variabel yaitu, kasus positif, kasus sembuh, dan kasus meninggal.

C. Pengolahan Data awal

Dari data yang sudah kita miliki, kemudian akan kita olah dengan menggunakan metode *K-Medoids Clustering*. Untuk mengetahui pengelompokan data sehingga data akan menampilkan hasil akhir dari pengelompokan *Clustering*.

D. Penerapan Algoritma *K-Medoids*

Algoritma *K-Medoids* biasa disebut sebagai *partitioning around medoids*, yang merupakan varian dari metode *K-Means*. Hal ini didasarkan pada penggunaan *medoids* bukan dari pengamatan *mean* yang dimiliki setiap *cluster*, yang bertujuan untuk mengurangi sensitivitas dari partisi yang dihasilkan sehubungan dengan nilai-nilai ekstrim yang ada pada *dataset*. Algoritma *K-Medoids* merupakan suatu algoritma yang mengatasi kelemahan Algoritma *KMeans* yang sensitif terhadap *outlier* karena objek dengan suatu nilai yang besar mungkin menyimpang dari distribusi data. Untuk perhitungan menggunakan algoritma *K-Medoids* dapat mengikuti langkah 1 sampai 6 seperti yang sudah dijelaskan sebelumnya. Dengan melakukan percobaan tiga *cluster* ($k=3$), sehingga didapatkan grafik perbandingan jumlah item 3 *cluster*.

Dalam Proses Perhitungan pada metode *KMedoids* berikut langkah-langkahnya sebagai berikut :

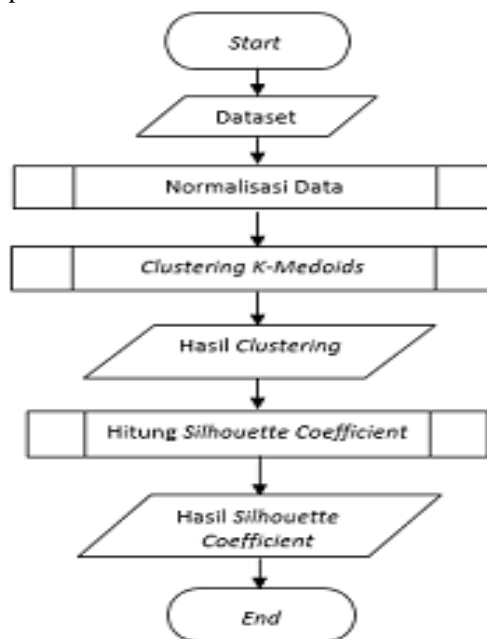
1. Melakukan normalisasi data yang akan digunakan dalam proses perhitungan *KMedoids*. Perhitungan normalisasi data menggunakan persamaan (1)
2. Menginisialisasikan pusat *cluster* secara acak, kemudian menghitung jarak data (objek) dengan pusat cluster menggunakan suatu *Euclidean Distance*. Dalam perhitungan *Euclidean Distance* menggunakan persamaan (2).
3. Menghitung total *distance* dari keseluruhan data di dalam *cluster*.
4. Menginisialisasikan pusat cluster baru secara acak kemudian menghitung jarak data (objek) dengan pusat cluster dengan *Euclidean Distance*. Perhitungan *Euclidean Distance* menggunakan persamaan (2).
5. Untuk menentukan atau menetapkan selisih total *distance* dengan mengurangi total *distance* baru – total *distance* lama.
6. Setelah itu untuk mendapatkan hasil *cluster* akhir, menghitung nilai *average dissimilarity* (a_i) untuk penilaian (evaluasi) dengan metode *silhouette coefficient*. Perhitungan $a(i)$ menggunakan persamaan (3)

7. Jika menetapkan dari perhitung nilai *lowest average dissimilarity* (b_i). Perhitungan $b(i)$ menggunakan persamaan (4).
8. Untuk menghitung nilai *silhouette coefficient* (s_i). Perhitungan $s(i)$ melakukan persamaan untuk mengenal mutu terbaik dari *cluster*.

III. HASIL DAN PEMBAHASAN

A. Perancangan

Pada proses perancangan yang akan dilakukan pada algoritma *K-Medoids*, dengan memasukkan jumlah data yang dibutuhkan dalam proses untuk menentukan jumlah clustering. Proses selanjutnya perhitungan seluruh di jaringan *K-medoids*. Pada hasil akhir akan menampilkan grafik/plot view nya didalam sebuah hasil perhitungan di rapid miner 5.3.



Gambar 2. Perancangan Perangkat Lunak

B. Data Awal

Data awal/sumber data yang digunakan dalam penelitian ini langsung dari kementerian kesehatan Indonesia. Data yang kami gunakan dalam penelitian yaitu alur grafik penyebaran covid-19 yang terjadi saat ini di Indonesia khususnya, dimana data ini menggunakan 34 provinsi yang ada di Indonesia. akan di kelompokkan di untuk menentukan *clustering* 1, 2, dan 3. Data penyebaran Covid-19 meliputi kasus positif, sembuh, dan meninggal. Berikut yaitu cara menentukan data / pengelompokan data.

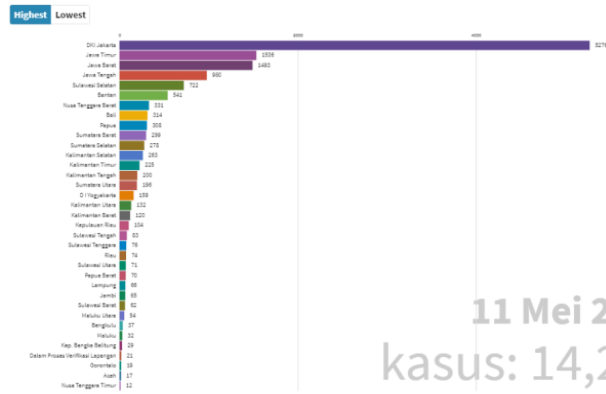
Dashboard Kasus COVID-19 di Indonesia :
2020-03-02 s/d sekarang



Gambar 3. Data Covid-19

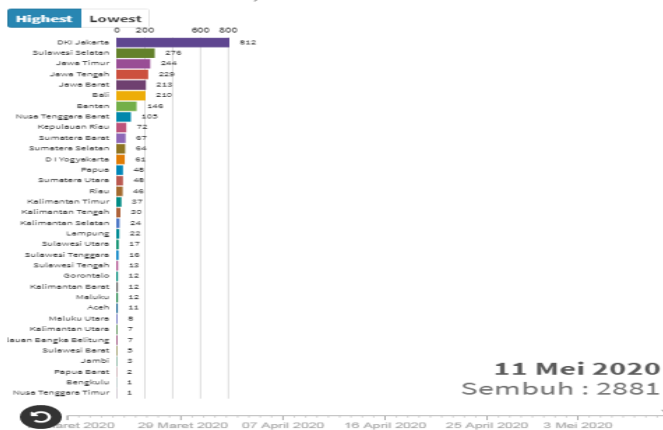
Pada data ini tingkat penyebaran yang terkonfirmasi positive covid ialah 14265 ribu jiwa, sembuh dari covid ada 2881 jiwa dan meninggal ada 991 diupdate sampai tanggal 11 mei 2020. Dari awal masuk corona virus (covid) ini di Indonesia mengalami peningkatan sejak awal teridentifikasi virus ini berbagai upaya pemerintahan Indonesia untuk mencegah penyebaran virus ini namun, kita juga harus patuh pada aturan pemerintah. Berikut adalah grafik peningkatan covid-19 di Indonesia.

Penyebaran Kasus Konfirmasi COVID-19 di Indonesia
Pusat Data dan Informasi, Kementerian Kesehatan



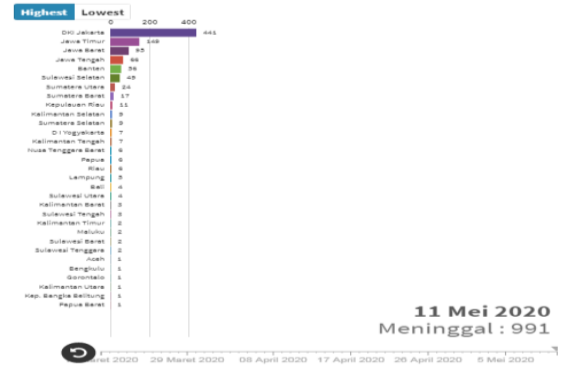
Gambar 4. Diagram Kasus Positif

Penyebaran Kasus Sembuh COVID-19 di Indonesia
Pusat Data dan Informasi, Kementerian Kesehatan



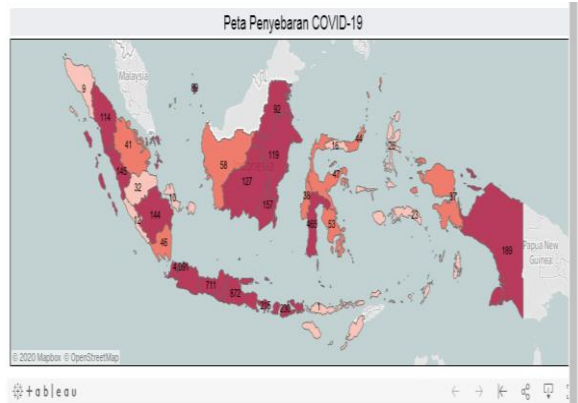
Gambar 5. Diagram Kasus Sembuh

Penyebaran Kasus Meninggal COVID-19 di Indonesia
Pusat Data dan Informasi, Kementerian Kesehatan



Gambar 6. Diagram Kasus meninggal

Dibawah ini adalah titik penyebaran Covid-19 dimana da 34 provinsi di Indonesia.

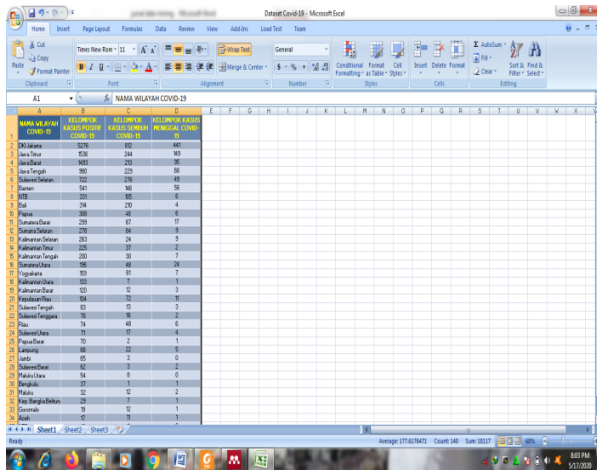


Gambar 7. Peta Penyebaran Covid di Indonesia

C. Hasil Analisa Pengujian

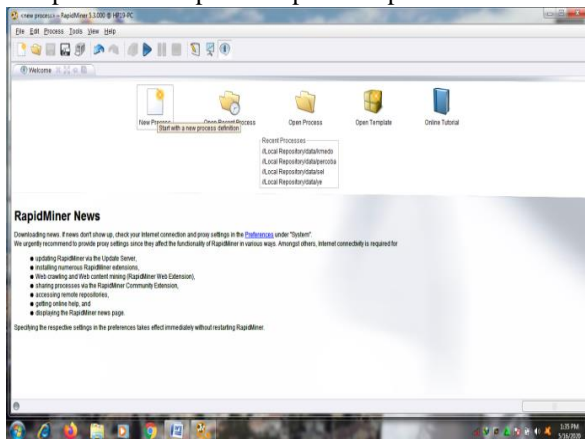
1. Masukan Proses Data

Pengumpulan input dataset excel di sejumlah wilayah yang terjangkit positif oleh covid, dalam proses data disini terdapat 2 tipe data masukan yaitu jumlah data 34 provinsi yang terjangkit/positive covid-19, dan jumlah cluster. Jumlah cluster disini ada 3 Clustering. Dimana jumlah setiap wilayah yang terpapar berbeda-beda.



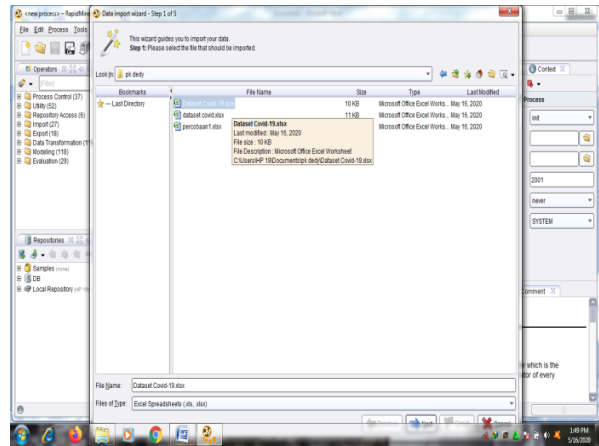
Gambar 8. Penyebaran di provinsi

Kemudian kita akan memproses dataset excel tadi menggunakan *software* Rapid Miner 5.3. untuk mengetahui penelompokan jumlah *cluster* yang didapat. klik *new* proses seperti tampilan dibawah ini.



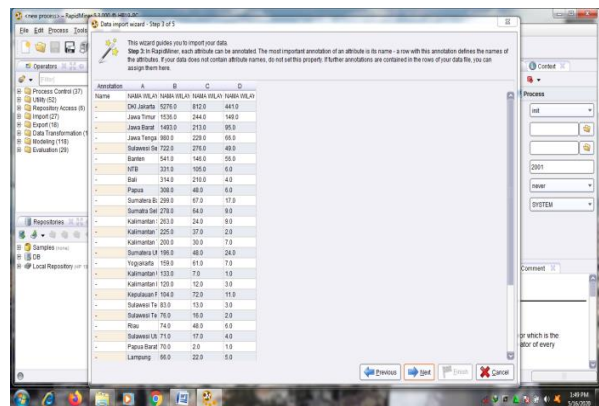
Gambar 9. proses pengolahan data

Lalu akan menampilkan tampilan proses seperti di bawah ini. Kemudian masukkan data yang akan di diproses ke lembar kerja. Dibagian kiri paling bawah klik symbol segitiga kebawah warna hitam klik (*import excel sheet*). Setelah kita klik import akan muncul tampilan seperti dibawah ini data set yang akan kita import dengan atribut yang sudah kita buat untuk mengetahui titik clustering yang kita butuhkan. klik data yang akan kita proses selanjutnya kemudian next.



Gambar 10. step 1 import data excel.

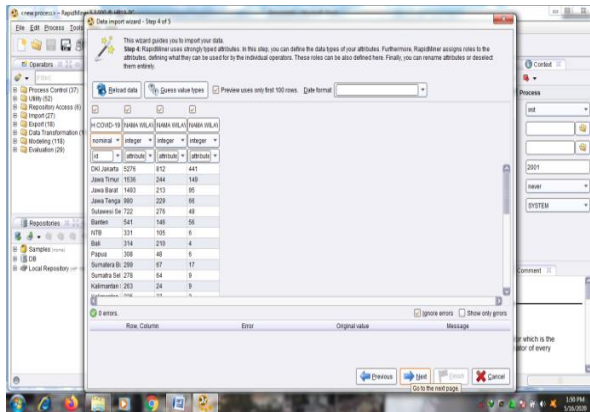
Kemudian tampil seperti gambar dibawah ini, disini tidak ada yang perlu di ubah klik next saja. Masih sama disini belum ada perubahan yang harus diubah, hanya saja variabelnya yang berubah. Perhatikan gambar dibawah ini kemudian di Next sampai proses step ke 3 tidak ada perubahan yang harus diubah Next.



Gambar 11. Step 3 Import Data Excel

Perhatikan pada Gambar dibawah ini :

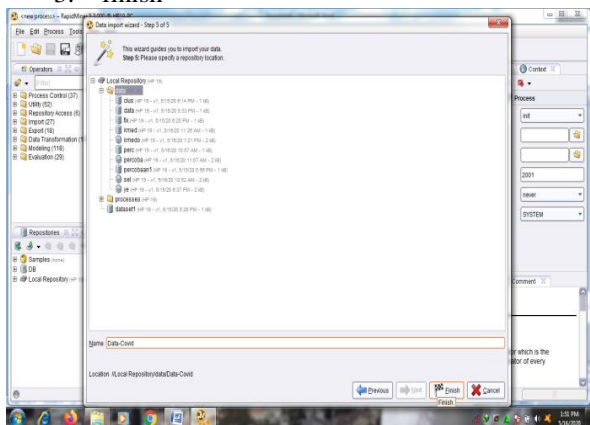
1. di bagian integer khusus wilayah *covid (binominal)* diubah menjadi (*nominal*)
2. dan di bagian (*atribut*) juga diganti/ dipilih menjadi (*id*). Kenapa diganti menjadi, karna untuk nama wilayah atau daerah mempunyai kunci tersendiri agar teridentifikasi nama wilayah tersebut.
3. Untuk dibagian yang lain biarkan saja tanpa ada perubahan sama sekali
4. Next



Gambar 12. Step 4 Import Data Excel

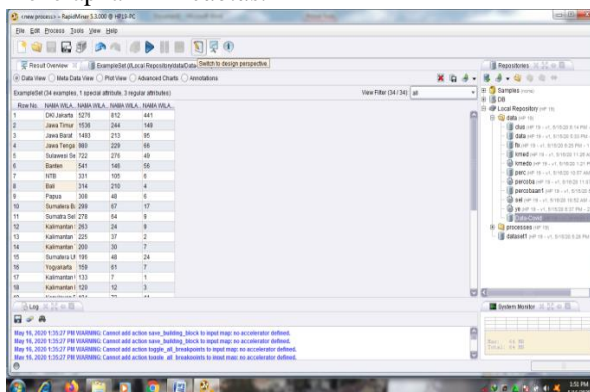
Kemudian simpan Data yang sudah kita ubah tadi perhatikan langkah-langkah pada gambar di bawah ini.

1. di (data)
2. buat file name yang akan disimpan
3. finish



Gambar 13. Step Import Data Wizard Selesai

Nah tampilan dibawah ini, tampilan dari *Data View* yang sudah siap kita proses selanjutnya untuk menerapkan *K-Medoids*.

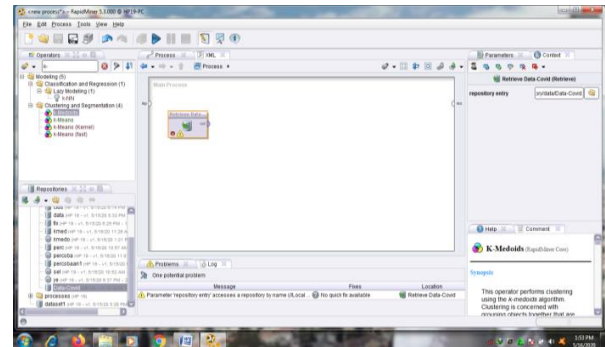


Gambar 14. Proses Import Berhasil

b. masukan penerapan Metode k-medoids

Disini kita akan menarik/drag data yang sudah kita simpan tadi dibagian bawah sebelah kiri. contoh filedata dalam penelitian yang digunakan

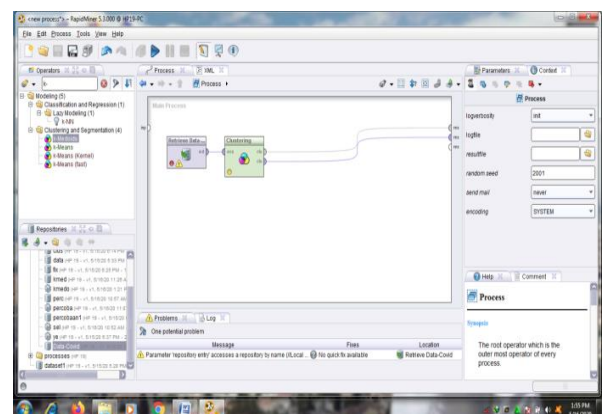
disini Data-Covid. Kemudian tarik Data-Covid tersebut ke lembar kerja yang ada ditengah seperti dibawah ini



Gambar 15. Input Data-Covid

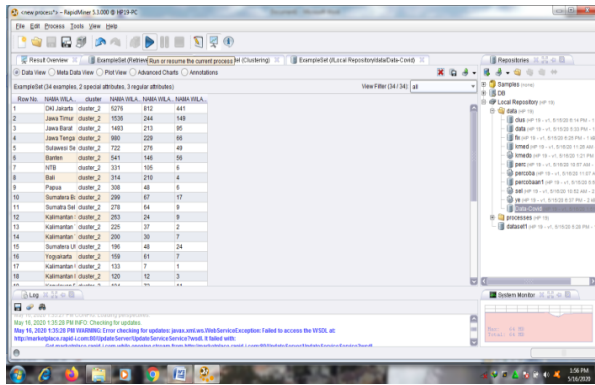
Tahap berikutnya disini yang paling penting menggunakan *K-Medoids*, karna penelitian ini menggunakan *K-Medoids*.

1. klik dibagian atas sebelah kiri cari/share (Modelling, Clustering and segmentation, K-Medoids)
2. di drag atau di tarik *K-Medoids* yang ada di tangan lembar kerja
3. kemudian hubungkan garis tersebut dengan benar.
4. Kemudian jalankan / run di atas dengan symbol segitiga warna biru



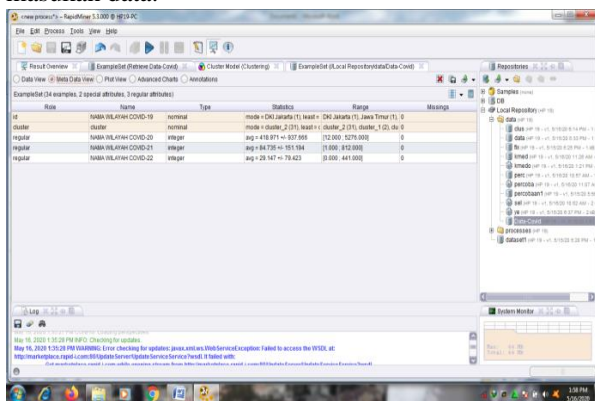
Gambar 16. clustering (K-medoids)

Proses run akan berjalan dengan baik jika tampilan data muncul pada saat run berhasil *Data View* seperti gambar dibawah ini.



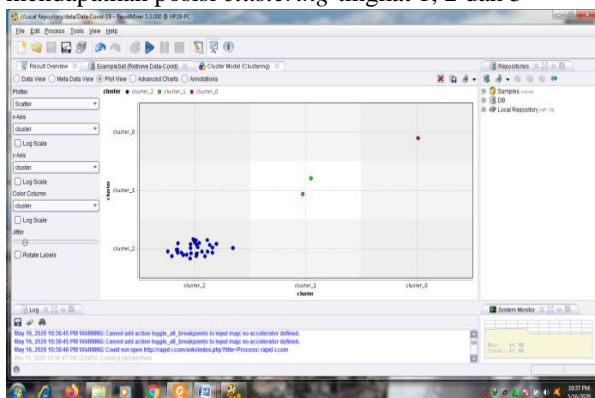
Gambar 17. Proses RUN

Tampilan Meta Data, disini kita bisa melihat jumlah dari cluster integer yang kita gunakan tadi di awal masukan data.



Gambar 18. Meta Data View

Dibagian *Plot View* kita dapat melihat garis dimana penentuan jumlah kelompok clustering yang ada di setiap daerah. Untuk tampilan plot view disebelah bagian kiri untuk x-Axis, y-Axis dan color custom pilihlah (*cluster*) semua yang akan terlihat pada gambar di bawah ini. Pada *Plot View* kita dapat mengetahui bagian-bagian kelompok mana saja yang mendapatkan posisi *clustering* tingkat 1, 2 dan 3



Gambar 19. Plot View dan hasil akhir

Pada gambar diatas ini adalah tampilan akhir dari penerapan K-Medoids untuk menentukan pengelompokan data *clustering*. Dari hasil pengujian

di atas, dapat disimpulkan bahwasanya *Cluster 0*, terdiri dari 1 wilayah. dan cluster 1 terdiri dari 2 wilayah, kemudian *cluster 2* terdiri dari 31 wilayah. Dari pengumpulan yang di dapatkan oleh rapitminer ini, dihitung dari *cluster* terkecil, menengah hingga *cluster* dengan nilai tertinggi yaitu C1,C2 dan C3. diketahui pola pemilihan penentuan pengelompokan penyebaran *covid-19* di berbagai wilayah di indonesia. K-Medoids merupakan metode Analitis *partisional clustering* yang bertujuan untuk mendapatkan suatu set *k-cluster* di antara data yang paling mendekati suatu objek dalam pengelompokan suatu data.. Hasil penelitian pengelompokan penyebaran *covid-19* baru menunjukkan bahwa masyarakat yang berasal dari berbagai wilayah di Indonesia.

IV. KESIMPULAN

Berdasarkan penelitian, implementasi dan pengujian, maka didapat Berdasarkan perhitungan yang telah dilakukan, algoritma *K-Medoids* dapat melakukan pengelompokan data covid-19 mana saja wilayah yang terinfeksi di wilayah masing-masing-masing dengan pengklasteran terbaik dilakukan dengan 3 *cluster*. Dari 34 *record* diperoleh 1 *record* pada *cluster* pertama, 2 *record* pada *cluster* kedua, 31 *record* pada *cluster* ketiga. Dan demikian untuk melakukan proses implementasi pada sistem dan analisis dapat menerapkan percobaan tersebut. Hasil dari percobaan juga dapat diimplementasikan dengan data yang besar dan atribut yang kompleks.

DAFTAR PUSTAKA

- [1] Atmaja, E. H. S. ‘Implementation of k-Medoids Clustering Algorithm to Cluster Crime Patterns in Yogyakarta’, *International Journal of Applied Sciences and Smart Technologies*, 1(1), pp. 33–44. doi: 10.24071/ijasst.v1i1.1859. 2019.
- [2] Juninda, T. and Andri, E. ‘Penerapan Algoritma K-Medoids untuk Pengelompokan Penyakit di Pekanbaru Riau’, (November), pp. 42–49. 2019.
- [3] Kementerian kesehatan Republik Indonesia <https://kementerian.kesehatan.go.id/data-covid-9>.
- [4] Marlina, D. *et al.* ‘Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak’, *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 4(2), p. 64. doi: 10.24014/coreit.v4i2.4498. 2018.
- [5] Metisen, B. M. and Sari, H. L. ‘ANALISIS CLUSTERING MENGGUNAKAN METODE K-MEANS DALAM PENGELOMPOKKAN PENJUALAN PRODUK PADA SWALAYAN

- FADHILA', 11(2), pp. 110–118. 2015.
- [6] Pramesti, D. F. *et al.* 'Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(9), pp. 723–732. doi: 10.1109/EUMC.2008.4751704. 2017.
- [7] Pulungan, N., Suhada, S. and Suhendro, D. 'Penerapan Algoritma K-Medoids Untuk Mengelompokkan Penduduk 15 Tahun Keatas Menurut Lapangan Pekerjaan Utama', *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 3(1), pp. 329–334. doi: 10.30865/komik.v3i1.1609. 2019.
- [8] Silitonga, D. A., Windarto, A. P. and Hartama, D. 'Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi', *Seminar Nasional Sains & Teknologi Informasi (SENSASI) SENSASI 2019 ISBN:*, pp. 313–318. 2019.
- [9] Wira, B., Budianto, A. E. and Wiguna, A. S. 'Implementasi Metode K-Medoids Clustering untuk Mengetahui Pola Pemilihan Program Studi', *Jurnal Terapan Sains & Teknologi*, 1(3), pp. 54–69. 2019.
- [10] Zayuka, H., Nasution, S. M. and Purwanto, Y. 'Perancangan Dan Analisis Clustering Data Menggunakan Metode K-Medoids Untuk Berita Berbahasa Inggris Design and Analysis of Data Clustering Using K-Medoids Method For English News', *e-Proceeding of Engineering* :, 4(2), pp. 2182–2190. 2017.