

**CLUSTERING DATABASE COVID 19 DI INDONESIA**

PROPOSAL SISTEM CERDAS

**Disusun Oleh:**

|  |  |
| --- | --- |
| **Nama** | **NIM** |
| **Abi Murti Eka S** | **18040084** |
| **Afrian Dwi Nanda** | **18040073** |
| **Moh. Ibnu Riski Kamal** | **18040045** |

**PROGRAM STUDI DIII TEKNIK KOMPUTER**

**POLITEKNIK HARAPAN BERSAMA TEGAL**

**2020**

**Abstrak**

*Abstrak*—Virus Corona (COVOD-19) adalah sebuah keluarga virus yang ditemukan pada manusia dan hewan yang menyebar antara manusia ke manusia melalui tetesan cairan dari mulut dan hidung saat orang yang terinfeksi sedang batuk atau bersin. Virus ini diketahui pertama kali muncul di pasar hewan dan makanan laut di Kota Wuhan.Virus Corona diketahui sudah mulai masuk ke Indonesia sejak Presiden Republik Indonesia (RI) Joko Widodo (Jokowi) mengumumkan dua warga negara Indonesia (WNI) yang terinfeksi Corona pada Senin, 2 Maret 2020. Dalam ilmu statistik, analisis kluster bertujuan untuk bertujuan untuk mengelompokkan *n* objek berdasarkan *p* variat yang memiliki kesamaan karakteristik diantara objek-objek yang akan diklasifikasikan ke dalam satu atau lebih *cluster* sehingga objek yang berada dalam satu *cluster* akan mempunyai kemiripan atau kesamaan karakter. Objek yang akan digunakan pada penelitian ini adalah data kasuss COVID-19 untuk 34 Provinsi di Indonesia tanggal 2 Maret 2020 sampai dengan 12 April 2020 dengan meggunakan dua metode Elbow dan metode Silhoutte*.* Data yang digunakan dalam praktikum ini merupakan data sekunder yang tersedia di *website* resmi Gugus Tugas Percepatan Penanganan COVID-19 dan Kawal COVID19 yaitu data mengenai Kasus Corna Virus (COVID-19) di Indonesia pada 2 Maret 2020 sampai dengan tanggal 12 April 2020 (42 hari). Berdasarkan analisis *cluster* metode Elbow dan metode Silhoutte untuk mencari jumlah cluster yang optimal, kedua metode tersebut menghasilkan *K*=3. Analisis *K-means clustering* dengan *K*=3 diperoleh kesimpulan hasil yang diperoleh ialah *good fit* karena memiliki variabilitas dalam cluster rendah dan antar cluster tinggi.

**1. PENDAHULUAN**

*World Health Organization* (WHO) menjelaskan bahwa *Coronaviruses* (Cov) adalah sistem pernafasan yang terinfeksi oleh virus[1]. Infeksi virus ini disebut Covid19. Virus Corona menyebabkan penyakit flu biasa sampai penyakit yang lebih parah seperti Sindrom Pernafasan Timur Tengah (MERS-CoV) dan Sindrom Pernafasan Akut Parah (SARS-CoV). Virus Corona adalah *zoonotic* yang artinya ditularkan antara hewan dan manusia. Berdasarkan Kementerian Kesehatan Indonesia, perkembangan kasus *Covid-19* di Wuhan berawal pada tanggal 30 Desember 2019 dimana Wuhan *Municipal Health Committee* mengeluarkan pernyataan “*urgent notice on the treatment of pneumonia of unknown cause*”. Penyebaran virus Corona ini sangat cepat bahkan sampai ke lintas negara. Sampai saat ini terdapat 212 negara yang mengkorfirmasi terkena virus Corona. Salah satunya di negara kita sendiri, Indonesia. Di Indonesia sendiri jumlah virus corona sudah mencapai 32 ribu lebih terjangkit virus ini. Penyebaran virus corona ini telah meluas ke berbagai daerah yang ada di Indonesia.

Covid-19 pertama masuk dilaporkan masuk ke Indonesia pada tanggal 2 Maret 2020 dengan jumlah 2 kasus. Kemudian dari data yang ada pada tanggal 31 Maret 2020 tercatat kasus virus ini meningkat sampai berjumlah 1.528 kasus terjangkit Covid-19. Tingkat moralitas di Indonesia mencapai angka

tertinggi di Asia Tenggara dengan angka sebesar 8.9 %. Dengan angka tingkat moralitas tertinggi di Asia Tenggara ini disebabkan karena angka kematian di Indonesia mencapai 136 jiwa. Selain itu jika ada yang terjangkit dan meninggal dunia pasti ada juga yang sembuh dari virus covid-19 ini, jumlah yang sembuh tercatat mencapai 81 orang.

Maksud kami mengambil judul ini karena melihat data- data banyaknya orang yang terjangkit, meninggal dunia, sampai sembuh yang terjadi di berbagai penjuru daerah di Indonesia. Untuk itu akan dibuat klasifikasi daerah dari zona merah, kuning, dan hijau yang berarti untuk zona merah sendiri berarti daerah itu adalah daerah bahaya dan daerah dengan jumlah terjangkit virus dengan jumlah cukup banyak. Untuk zona kuning berarti zona itu masuk di daerah dengan zona siaga, maksudnya adalah di daerah tersebut masuk daerah bisa dibilang dengan korban terjangkit virus Covid-19 dengan jumlah yang tidak banyak namun tidak juga sedikit. Yang terakhir zona hijau yang berarti zona aman, maksudnya di zona ini daerah yang terjangkit virus ini datanya paling sedikit terjangkit. Pengklasifikasian tersebut menggunakan metode *Clusstering (K-Means).*

**2. METODE PENELITIAN**

**Data Mining**

Menganalisis data yang prespektif dan menyimpulkan menjadi informasi penting yang dapat digunakan untuk meningkatkan keuntungan, biaya pengeluaran diperkecil, atau juga bisa keduanya merupakan pengertian dari *data mining*[2]. *Data mining* adalah analisis pada data untuk mencari hubungan yang jelas dan menyimpulkan belum diketahui sebelumnya dengan cara terkini agar mudah dipahami dan bermanfaat[2].

**Tahap-Tahapan dalam Data Mining**

Ada tahap-tahapan dalam data mining, yang bersifat interaktif, pengguna langsung terlibat atau dengan perantara *knowledge base*. Berikut tahap-tahap dari *data mining*[2]:

1. Pembersihan data (*Data Cleaning*)

Pembersihan data merupakan proses menghilang-kan noise dan data yang tidak konsisten atau data tidak relevan.

1. Integrasi data (*Data Integration*)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru.

1. Seleksi data (*Data Selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.

1. Transformasi data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam Data Mining.

1. Proses Mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan Data Mining.

1. Evaluasi pola (*Pattern Evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam knowledge based yang ditemukan.

1. Presentasi pengetahuan (*Knowledge Presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

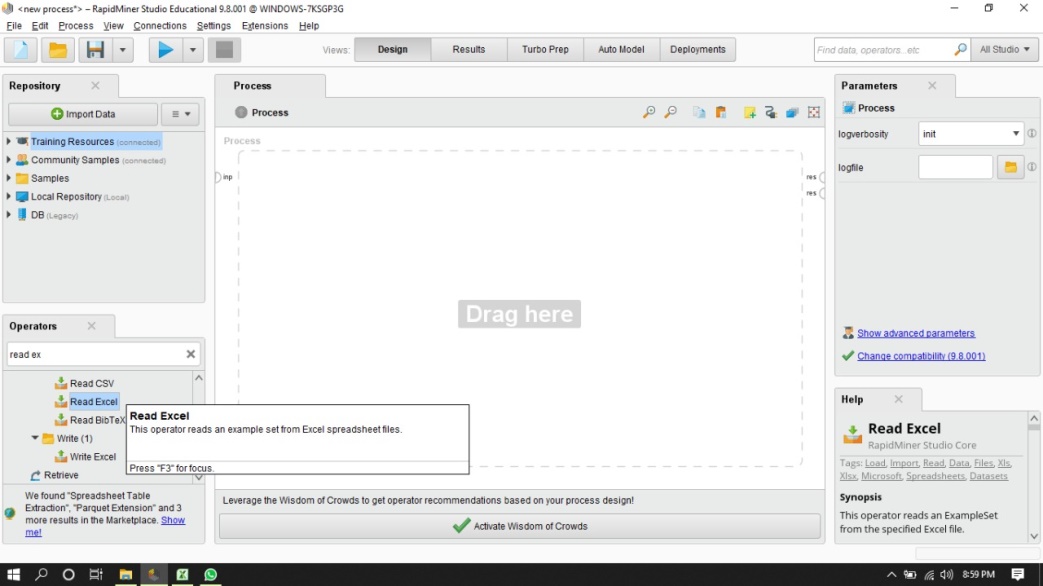
**Metode *Clustering***

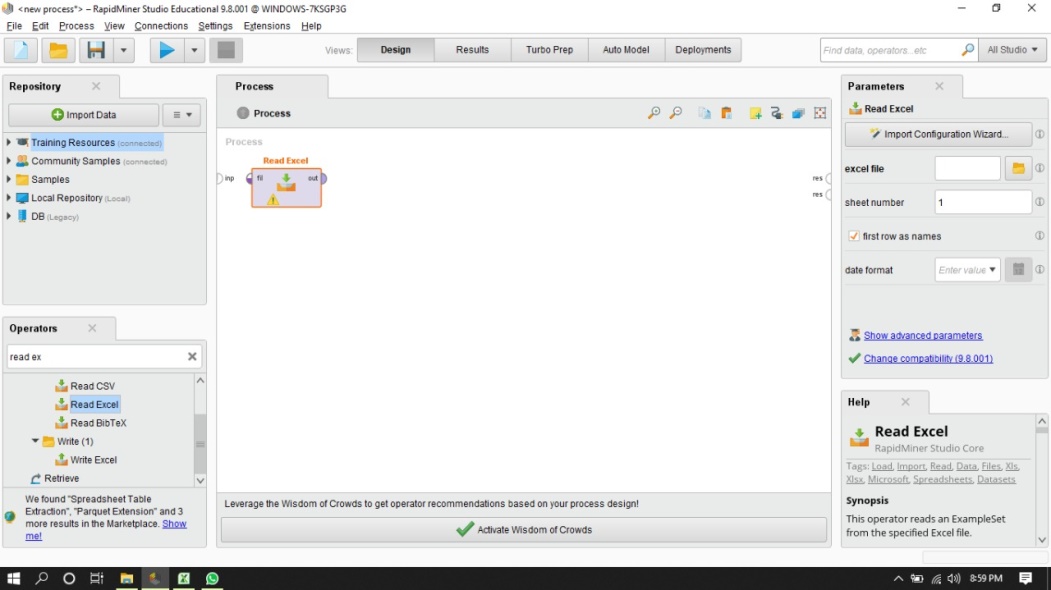
*Clustering* adalah memisahkan data ke dalam kelompok-kelompok yang memiliki obyek dengan karakteristik yang serupa[3]. Clustering ini merupakan awalan dari sebuah data mining.

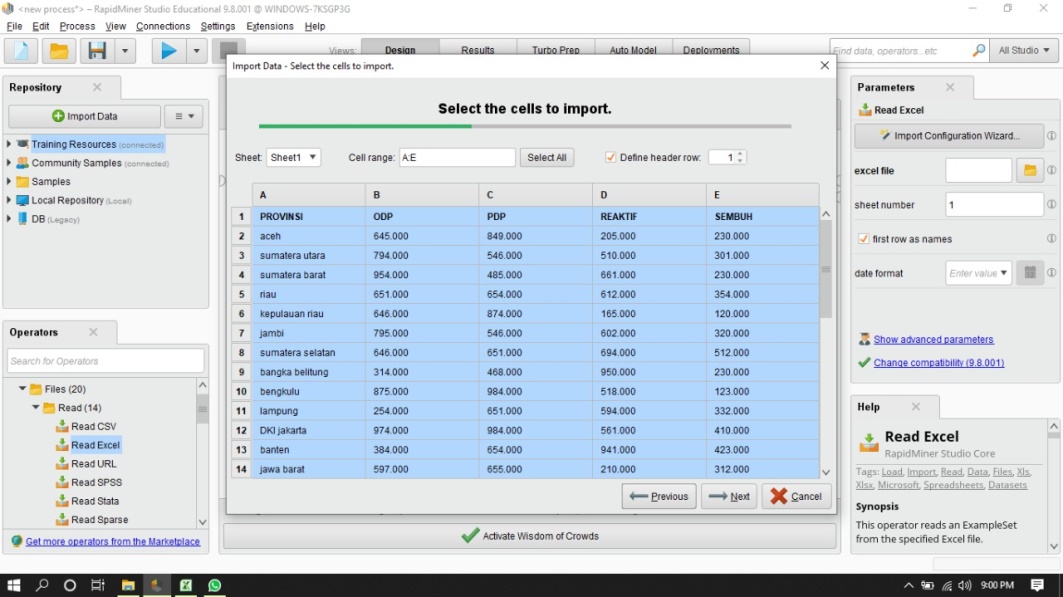
Dengan menggunakan clustering ini, dapat menklasifikasikan pasien Covid-19 dari tiap provinsi di Indonesia. Adapun tujuan dari cluster ini, yaitu dapat menentukan status pada setiap provinsi.

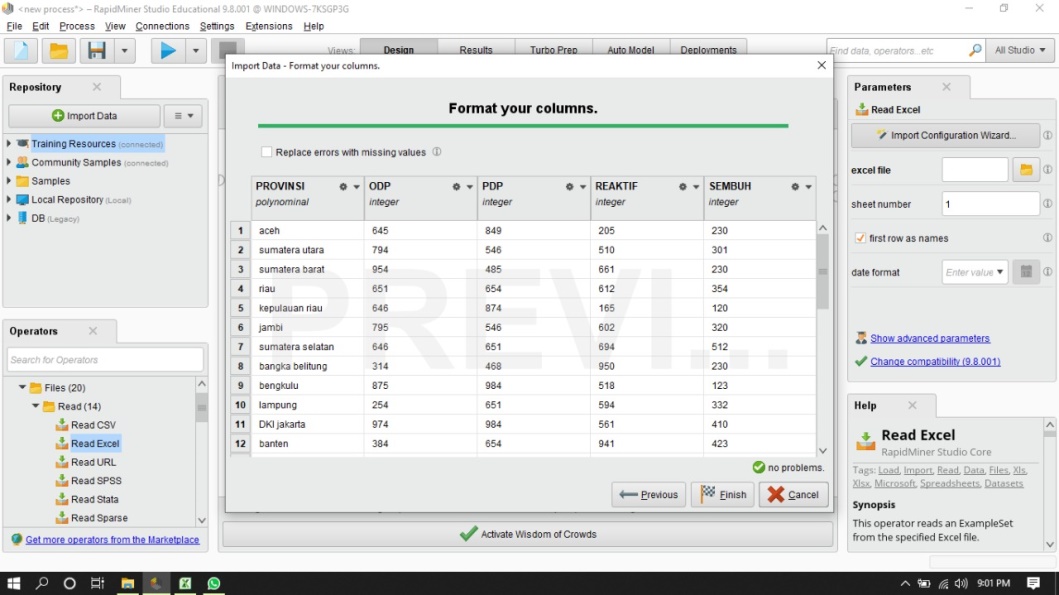
**HASIL DAN PEMBAHASAN**

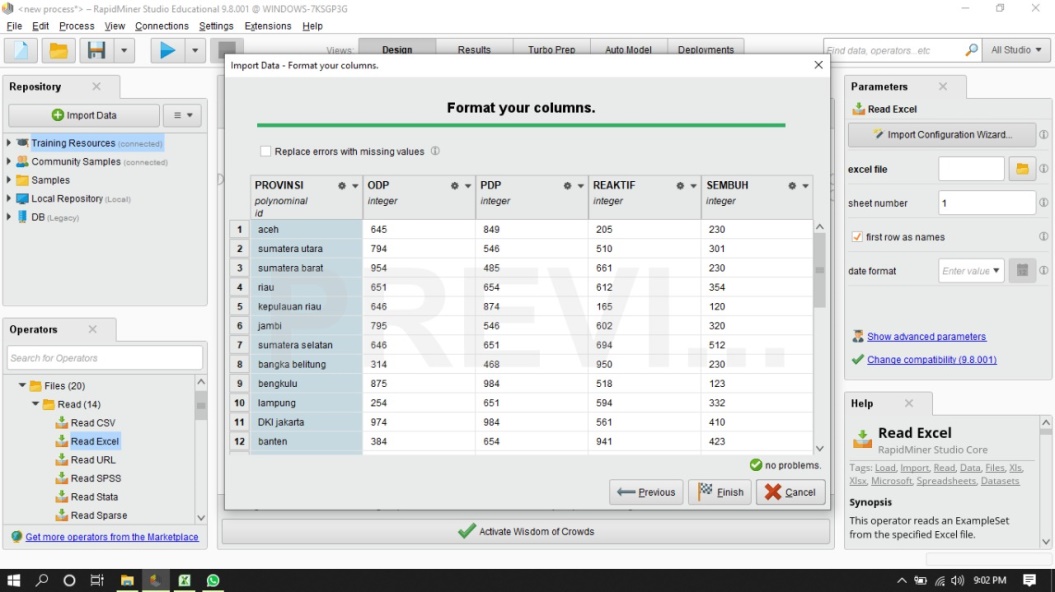
**Menggunakan rapitminer**

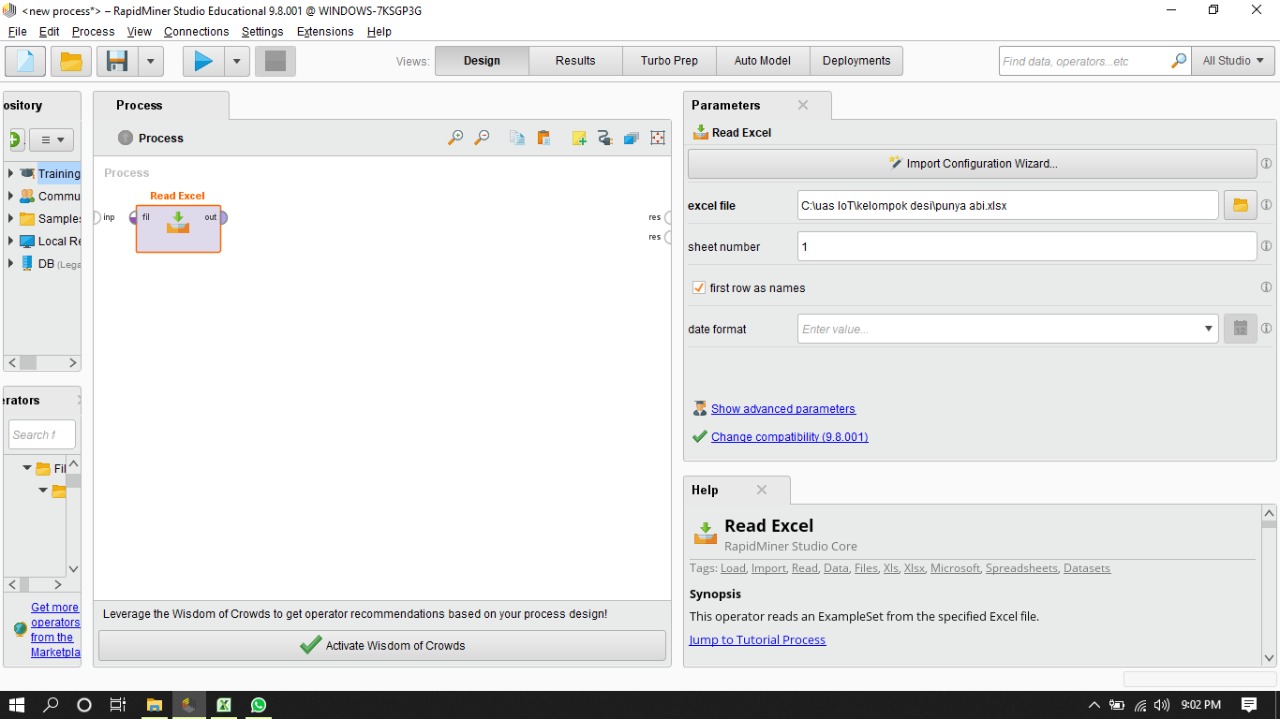
cari read excel di operations untuk memasukan dataset****

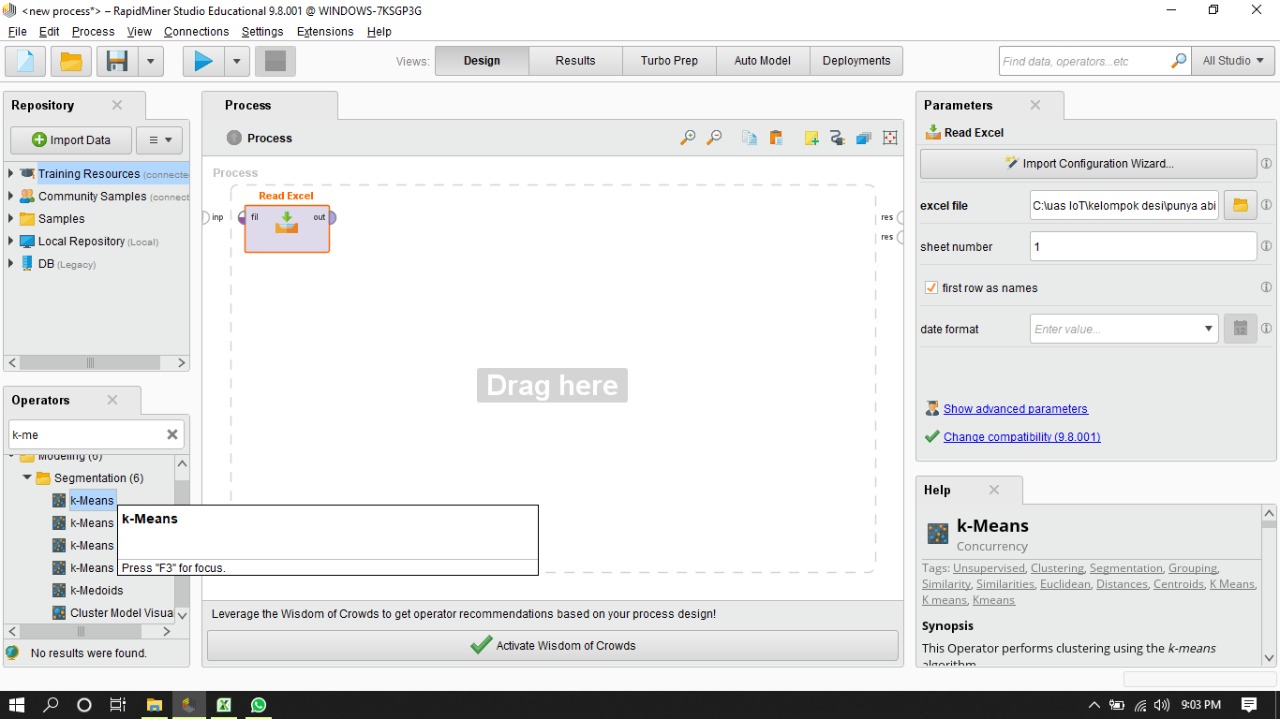
drag atau double klik sampai masuk ke lembar main proses****

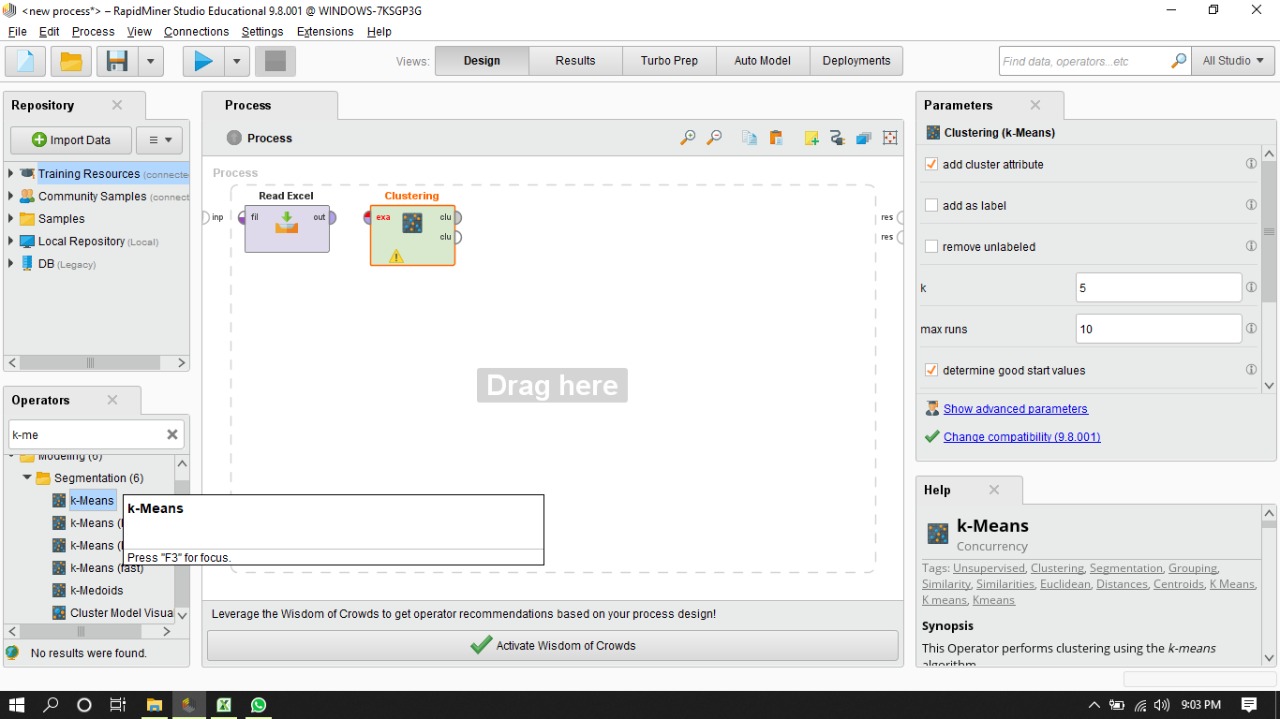
import database supaya masuk ke lembar kerja, atur jumlah colomns untuk actions selanjutnya****

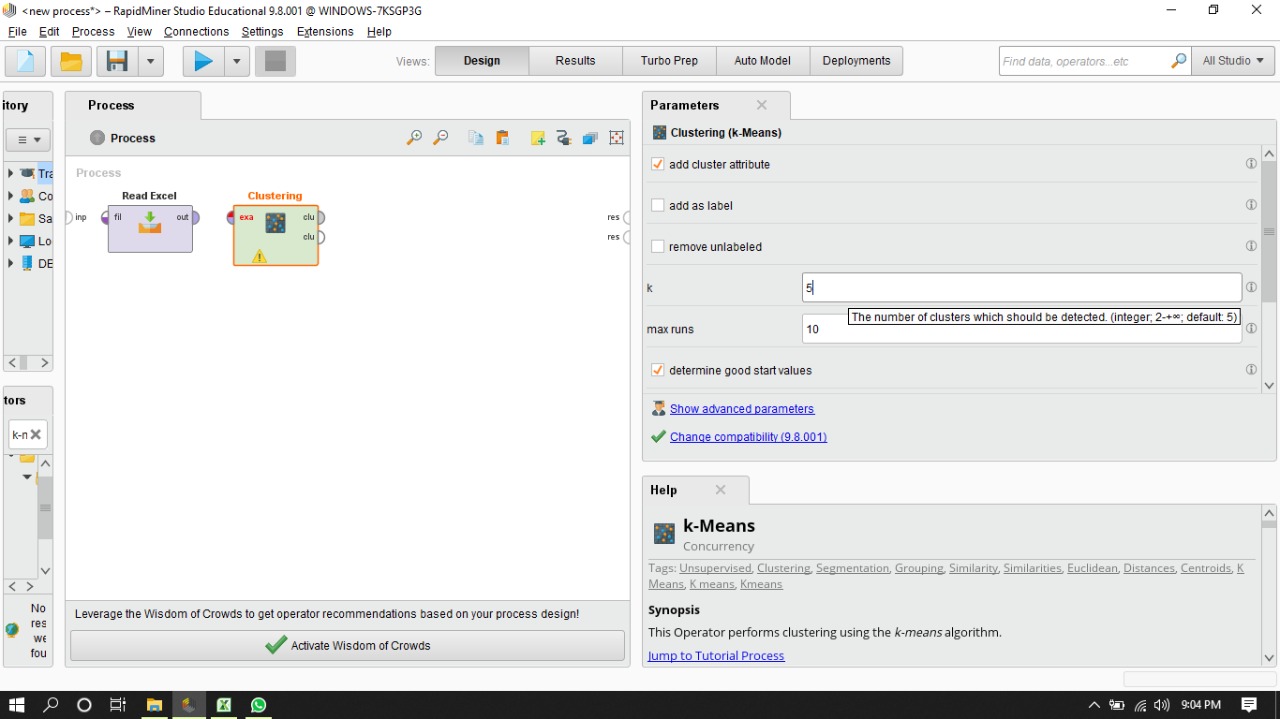
atur format colomns sesuai kebutuhan****

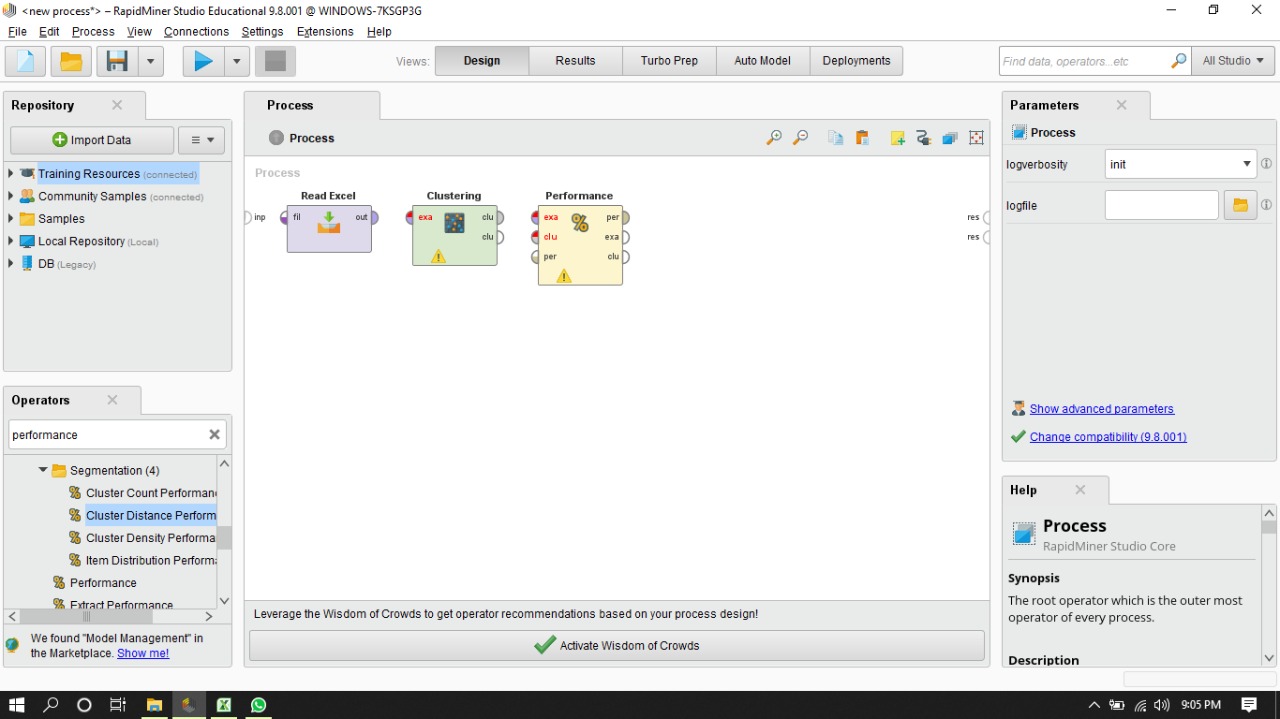
rubah format provinsi menggunakan format id untuk acuan data****

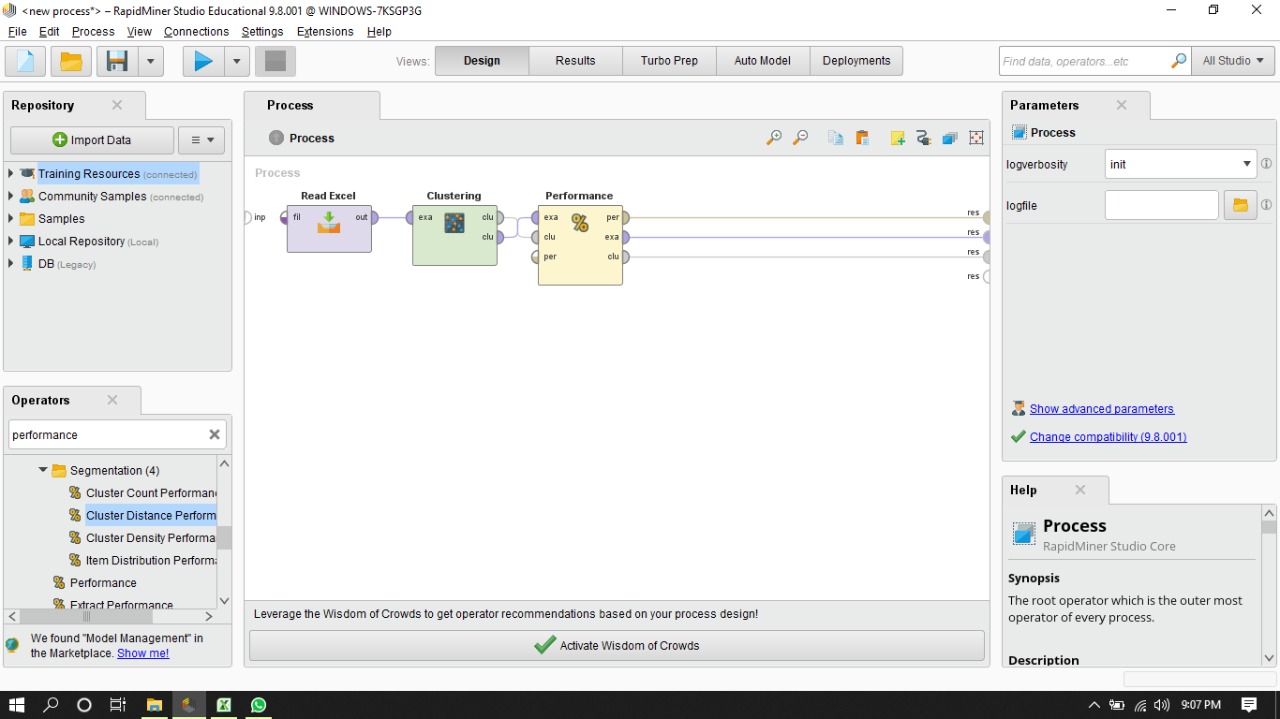
database berhasil terimport****

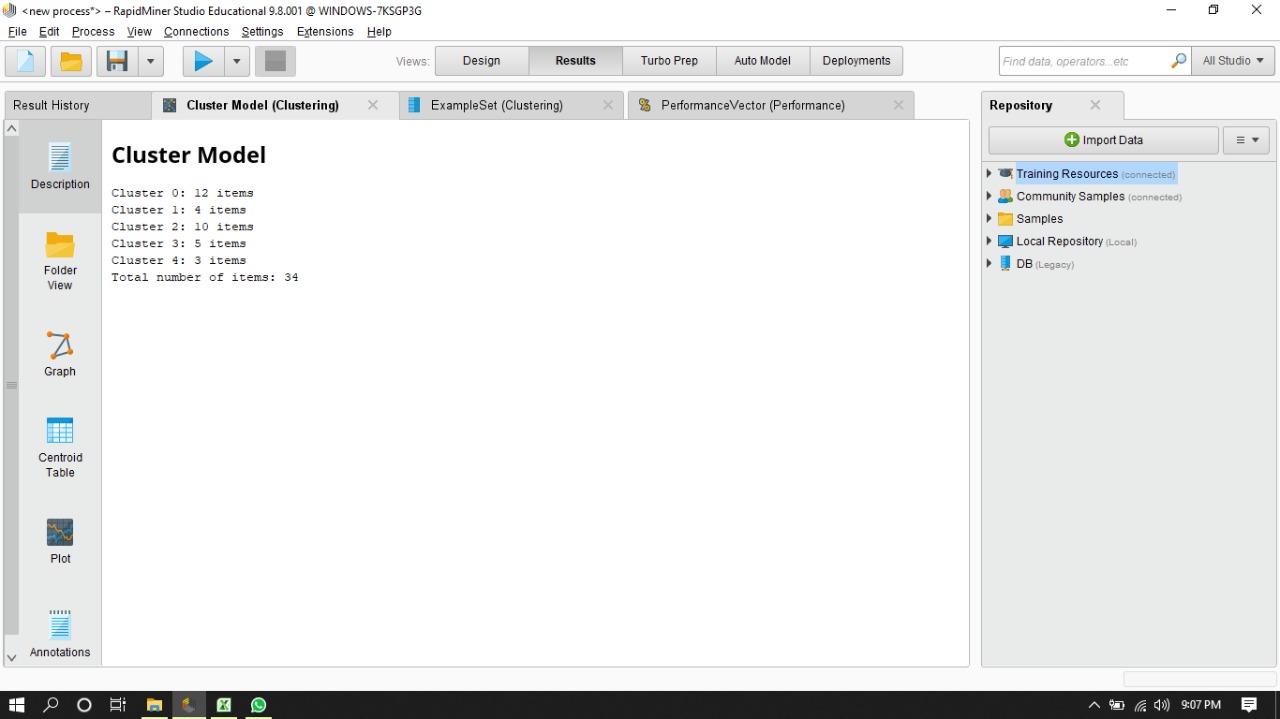
kemudian cari metode k-means pada operations****

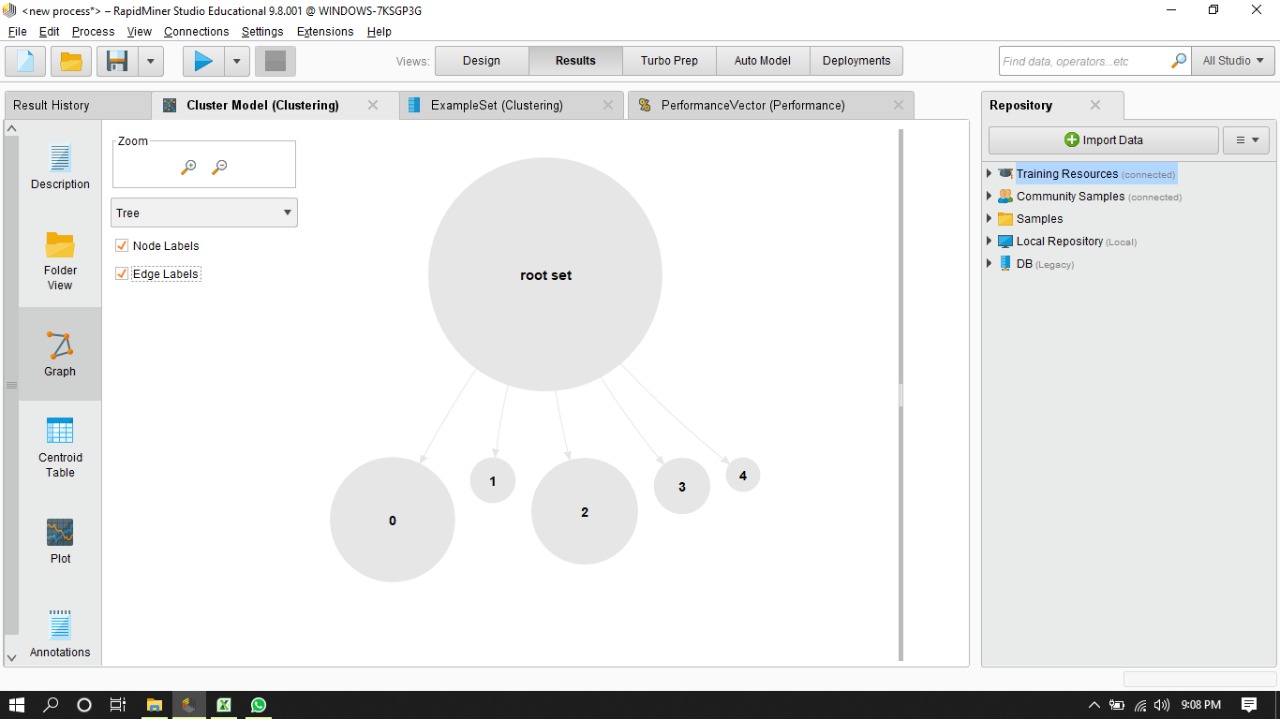
drags atau double klik sampai k-means masuk ke main proses****

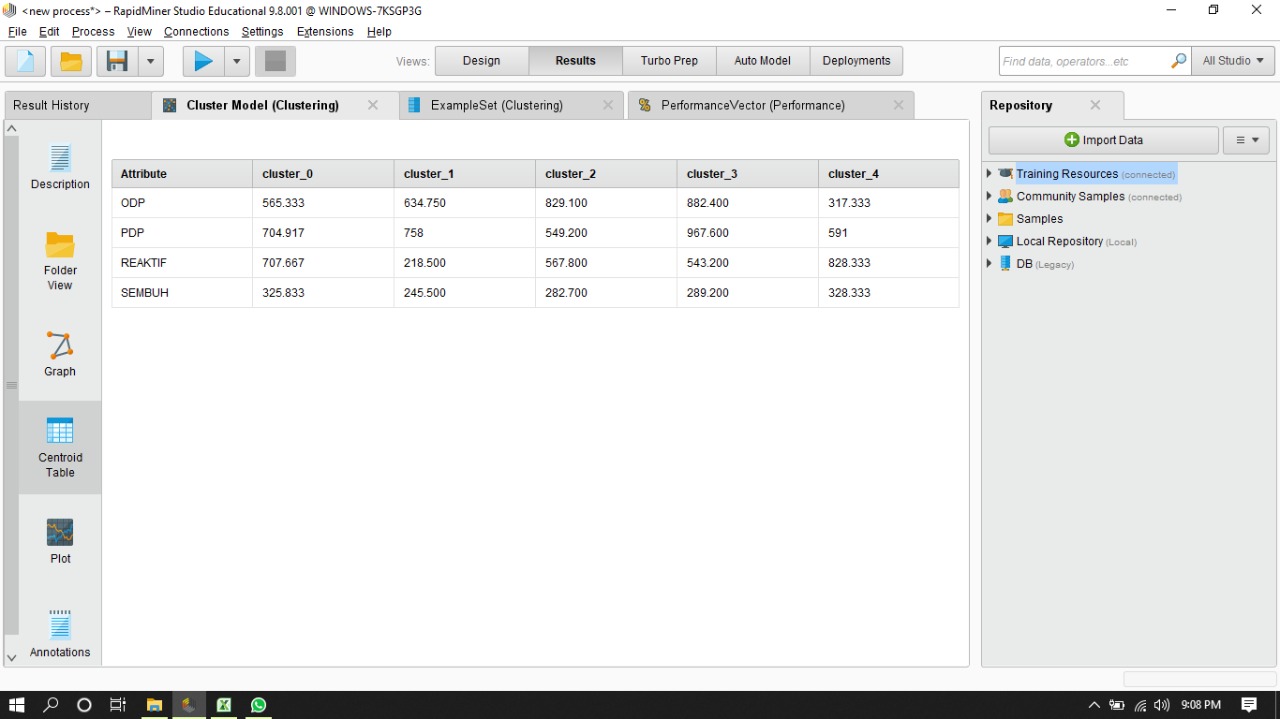
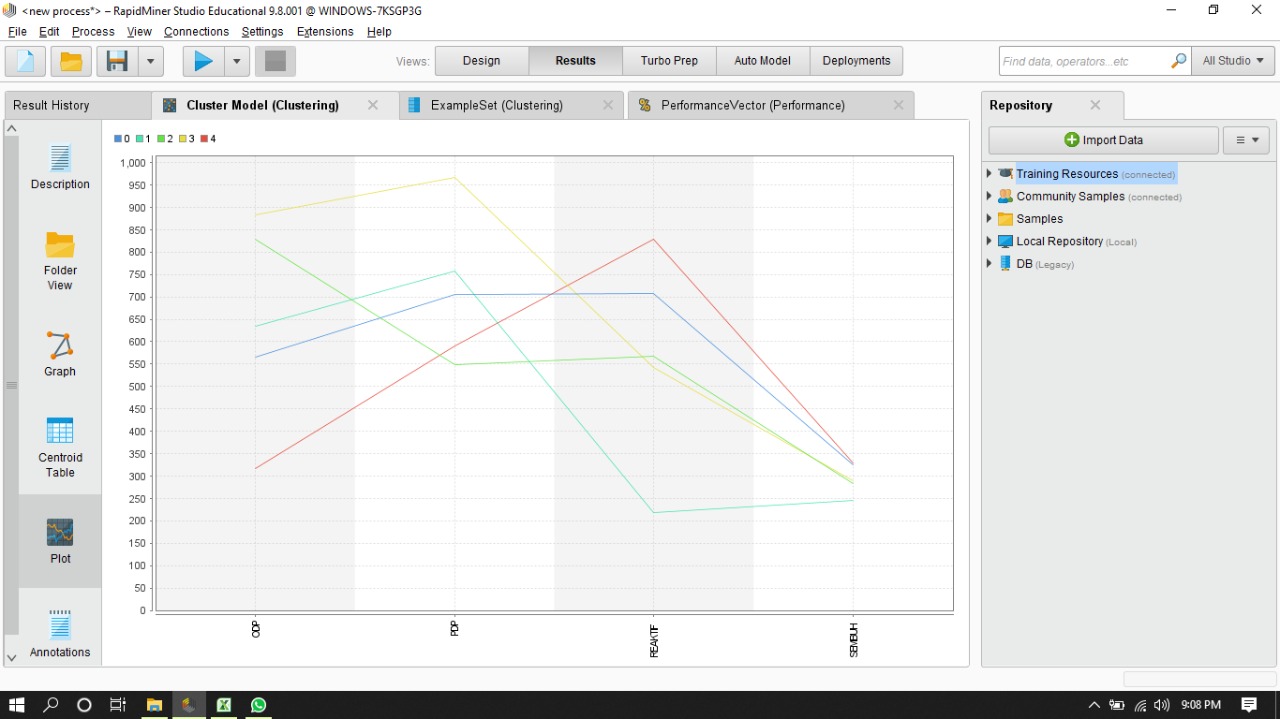
atur jumlah columns yang akan di clustering****

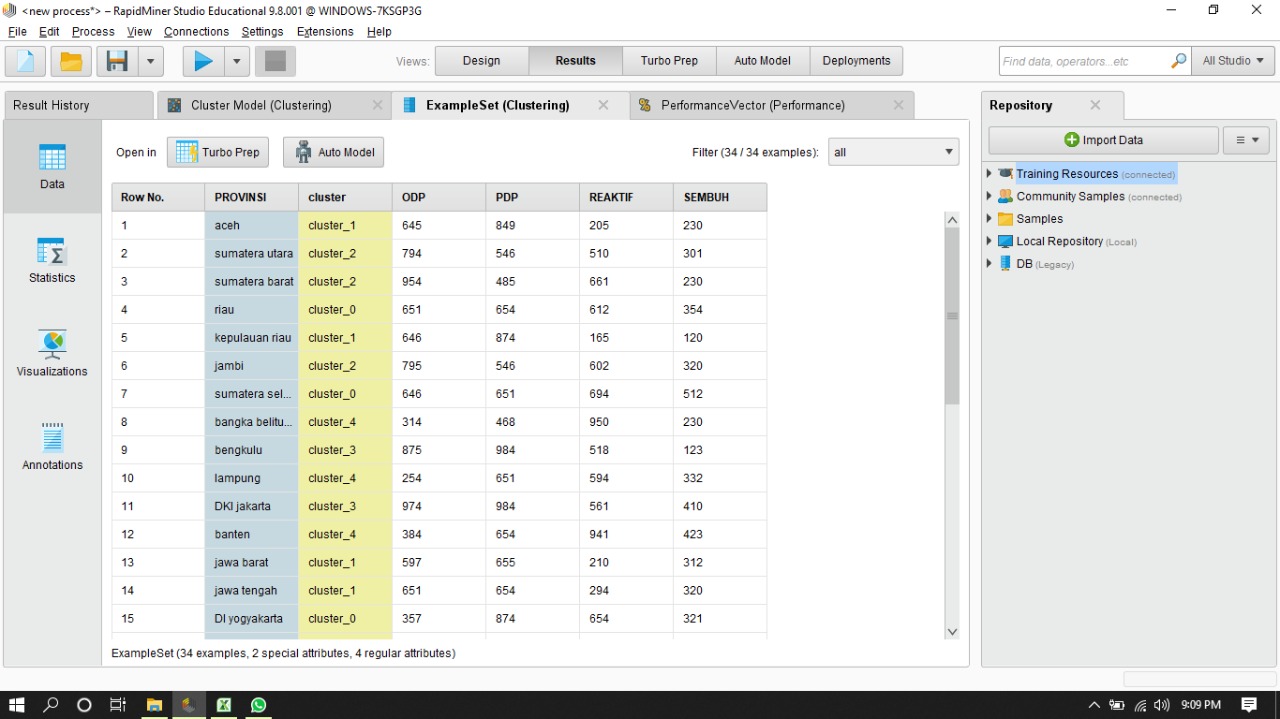
masukan performance distance agar dataset hasil clustering bisa ditampilkan dalam format exce****

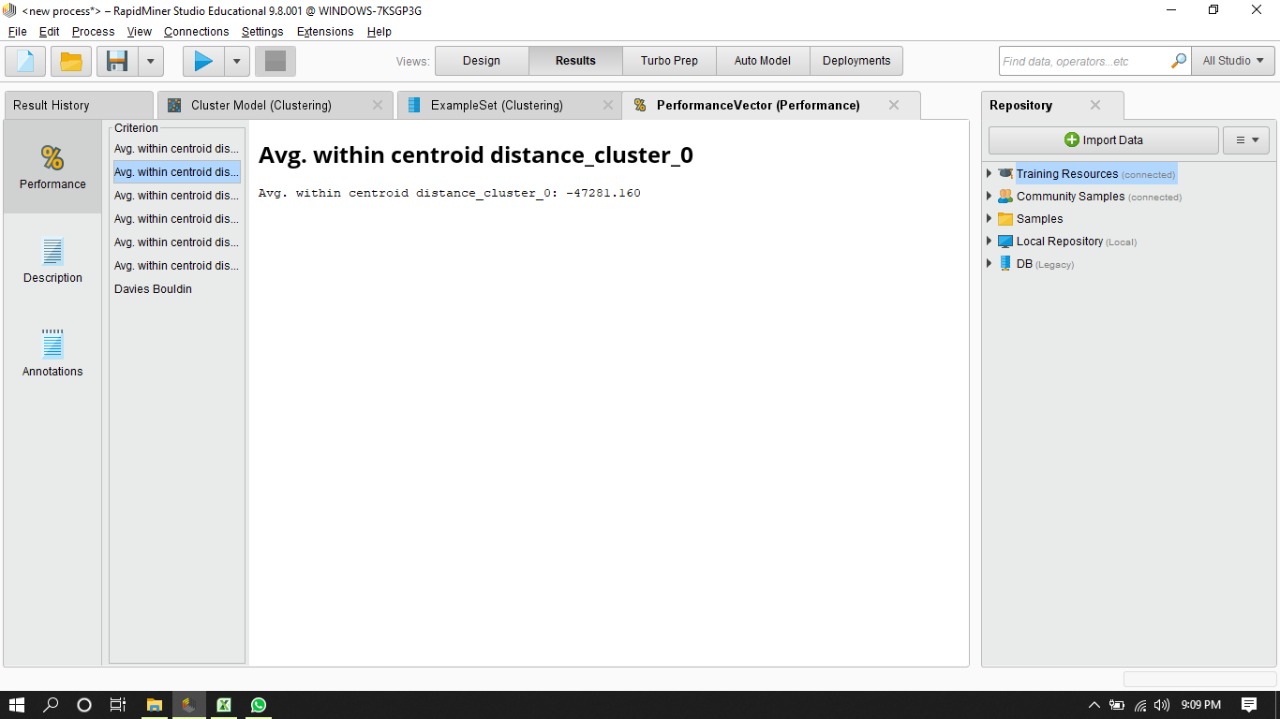
hubungkan atau koneksikan semua yang ada pada lembar kerja ****

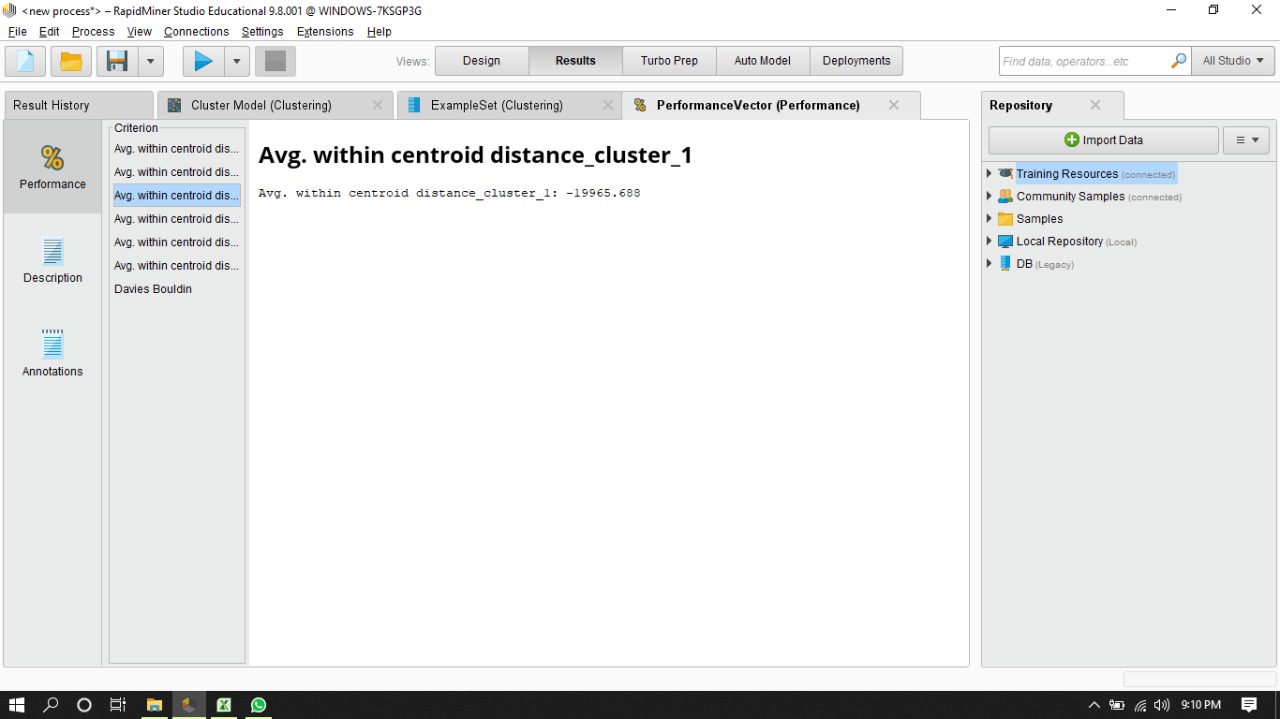
maka akan muncul hasil model cluster sebagai berikut****

grapich pengelompokan data****

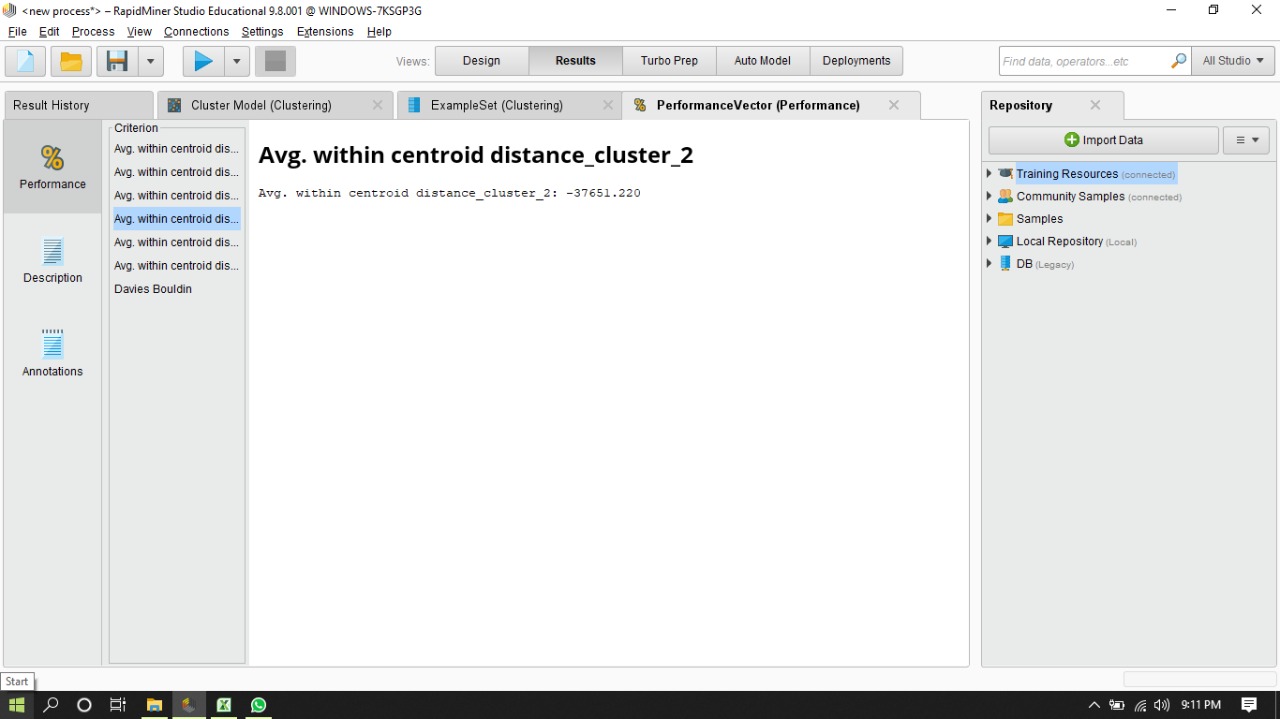
hasil pengelompokan data****

tampilan dataset yang sudah jadi dan sudah digolongkan atau diklasifikasi sesuai clusternya

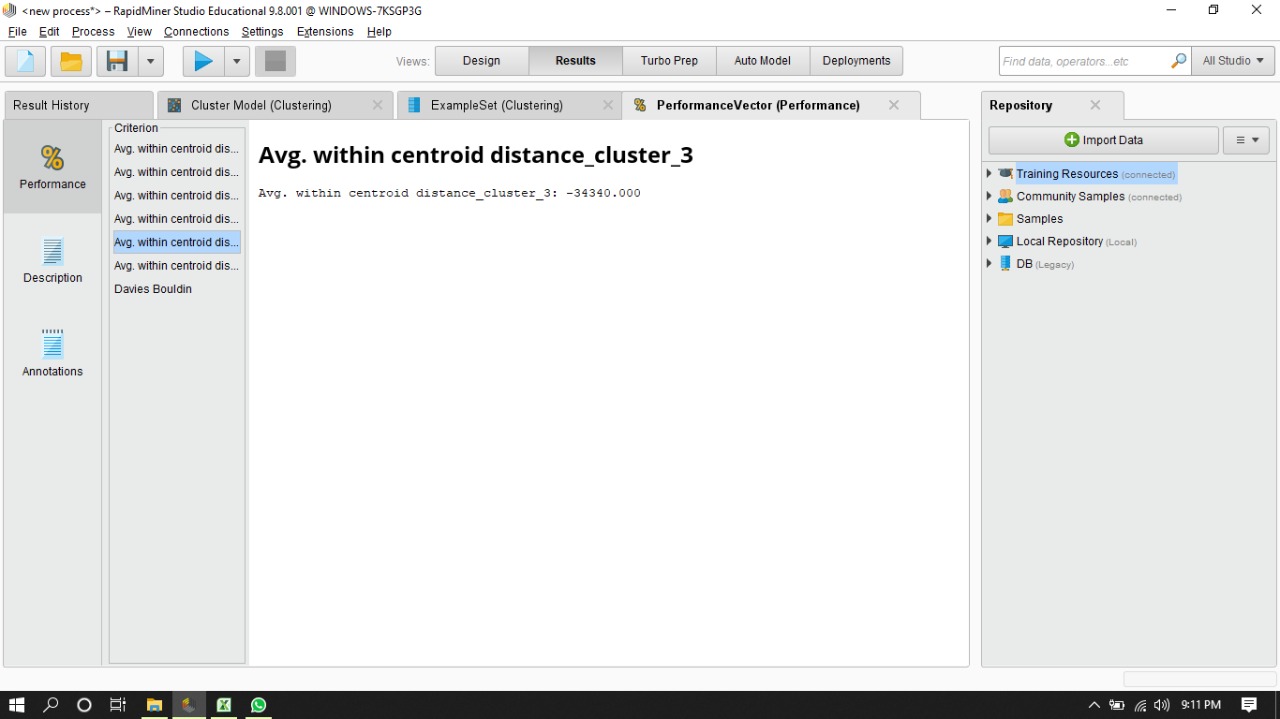
Average withing centroid distance pada cluster 0****

average within centroid distance pada cluster 1****

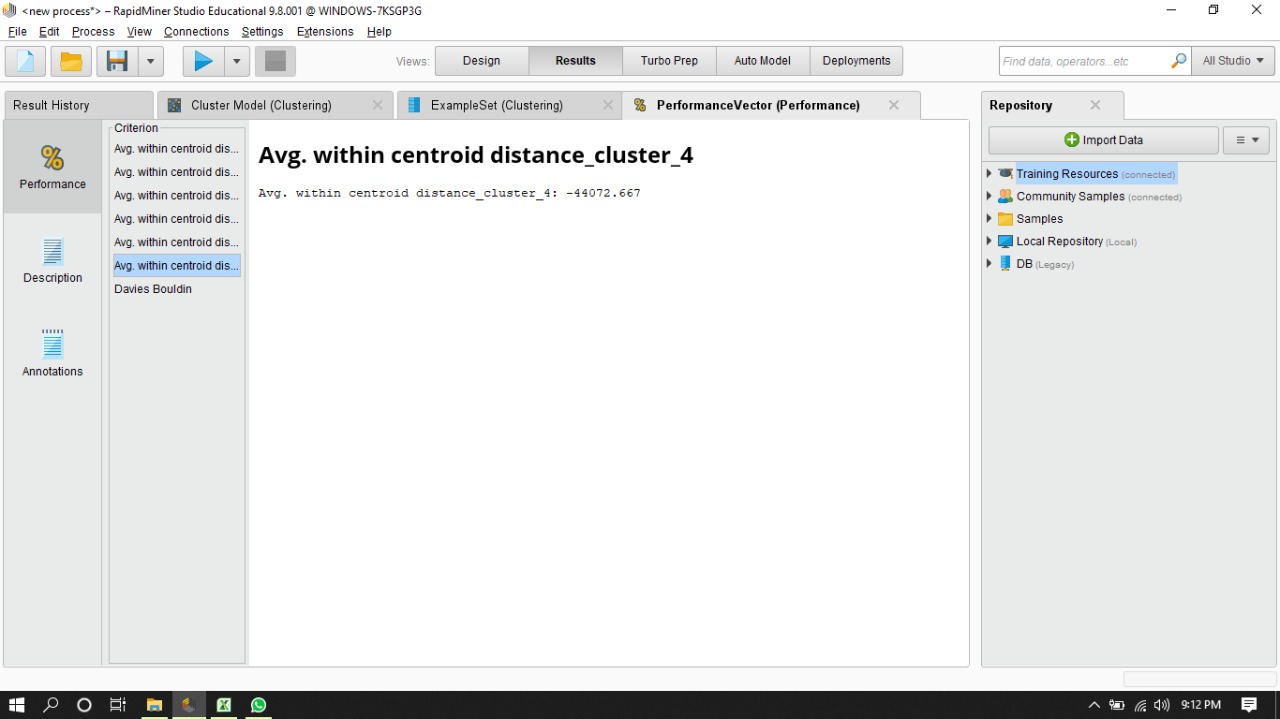
Cluster2

****

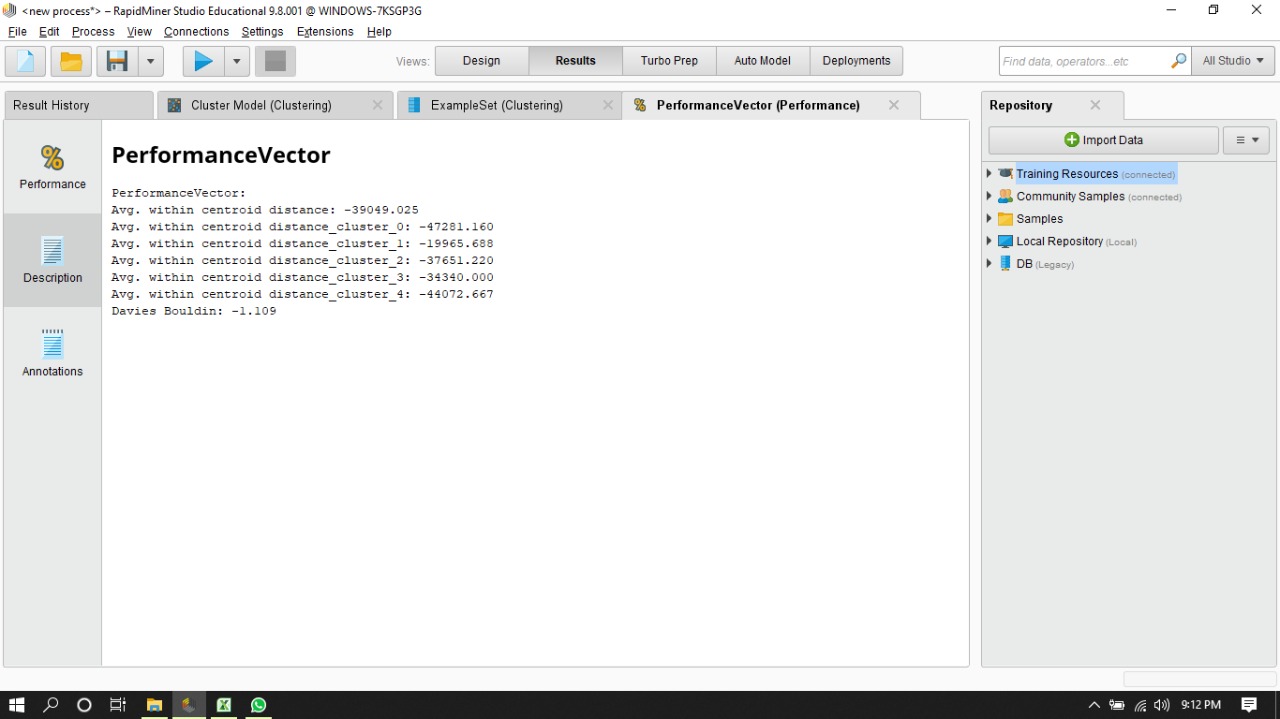
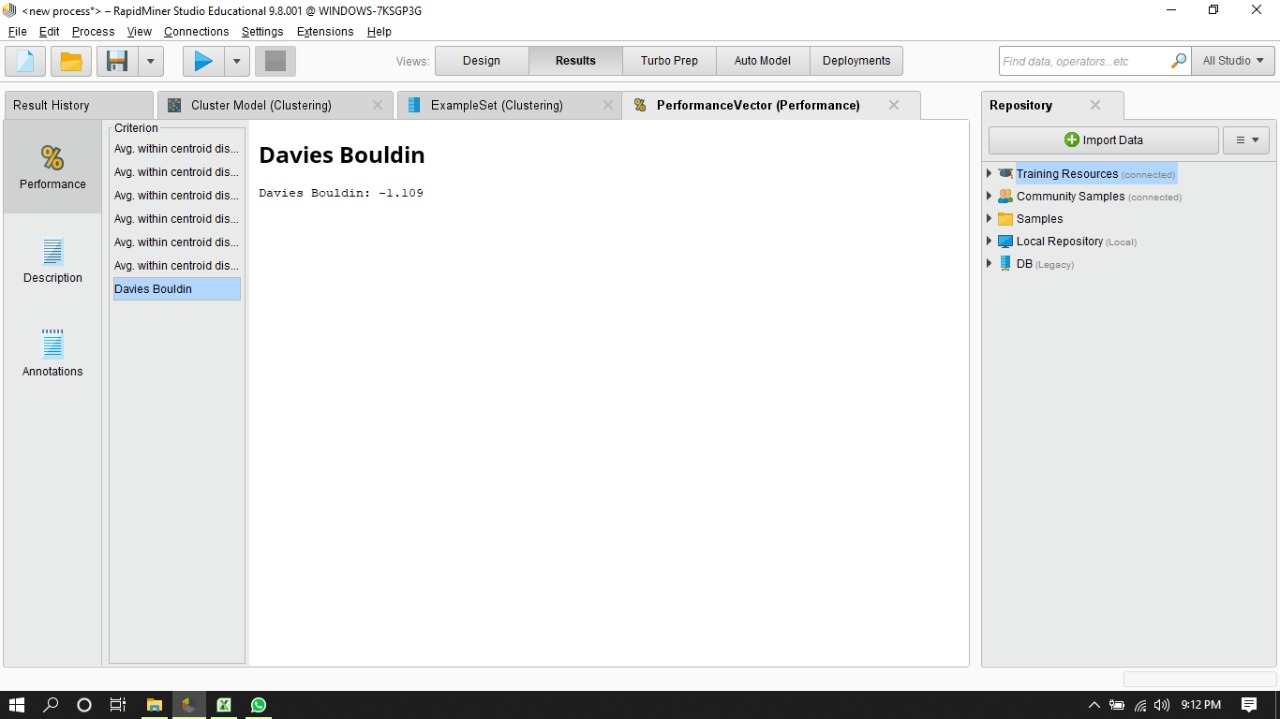
Cluster3

****

Cluster4

****

Hasil Keseluruhan

****

**SIMPULAN**

Berdasarkan pengujian yang dilakukan menggunakan *Algoritma K-Means Clustering*, menyatakan bahwa:

1. Pusat cluster yang diperoleh untuk cluster pertama atau zona hijau berada pada provinsi Aceh, Bali, Bangka Belitung, Bengkulu, DIY, Jambi, Kalimantan Barat, Kalimantan Timur, Kalimantan Tengah, Kalimantan Utara, Kepulauan Riau, NTB, Sumatera Barat, Sumatera Utara, Sulawesi Tenggara,Sulawesi Tengah, Lampung, Riau, Maluku, Maluku Utara, Papua, Papua Barat, NTT, Gorontalo.

2. Pusat cluster yang diperoleh dari cluster kedua atau zona merah berada pada provinsi DKI Jakarta dan Jawa Timur.

3. Pusat cluster yang diperoleh dari cluster ketiga atau zona kuning berada pada provinsi Banten, Jawa Barat, Jawa Tengah, Kalimantan Selatan, Sumatera Selatan, Sulawesi Selatan.

4. SARAN

Adapun saran yang dapat diberikan mengenai penelitian tentang menentukan status provinsi di Indinonesia, dapat menentukan status provinsi dengan berbagai metode tidak hanya dengan *clustering*.

**DAFTAR PUSTAKA**

[1] Pertanyaan jawaban terkait COVID-19 untuk publik. (n.d.).

[2] Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. 2(3), 207–217.

[3] Method, I., Based, K. C., Value, S., Interface, W., Study, C., & Magelang, I. U. M. M. (2015). Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. 18(1), 76–82.