



Penerapan Data Mining untuk clustering data Penduduk yang Terdampak Covid-19 menggunakan Algoritma K-means

Syahril Rizal¹, Rindi qusnul Khotimah^{2*}

^{1,2}Program Studi Teknik Informatika, Fakultas ilmu komputer, Universitas Bina Darma Palembang
Email: syahril.rizal@binadarma.ac.id, rindiqusnul28@gmail.com

Abstrak

Sejak desember 2019 penyebaran virus covid-19 mulai terjadi di berbagai belahan dunia, tak terkecuali di Indonesia khususnya di provinsi Sumatra selatan virus corona terus meningkat setiap harinya. Virus tersebut sangat berdampak pada perekonomian di Sumatra selatan, jumlah pengangguran yang terus meningkat setiap harinya dan pengurangan pekerja yang terus bertambah, penelitian ini bertujuan untuk mengetahui berapa banyak pekerja yang terdampak covid-19 dengan menggunakan data jenis pekerjaan dari tahun 2019 sampai dengan 2021 yang ada di Sumatra selatan, data tersebut ada di Badan Pusat Statistik (BPS), maka dari itu penulis menggunakan metode clustering algoritma K-Means, teknik pengolahan data menggunakan tahapan KDD Knowledge Discovery in Database, hasil yang di dapatkan dalam penelitian ini adalah pada tahun 2019 jumlah pekerja masih normal, pada tahun 2020 mengalami sedikit pengurangan di tahun 2021 mengalami penurunan sebanyak 3% dari 3 tahun terakhir.

Kata Kunci: *covid-19, algoritma K-means, badan pusat statistik*

Abstract

Since December 2019 the spread of the Covid-19 virus has begun to occur in various parts of the world, including in Indonesia, especially in the province of South Sumatra, the corona virus continues to increase every day. The virus has greatly impacted the economy in South Sumatra, the number of unemployed continues to increase every day and the reduction in workers continues to grow, this study aims to find out how many workers are affected by COVID-19 by using data on types of work from 2019 to 2021. in South Sumatra, the data is in the Central Statistics Agency (BPS), therefore the author uses the K-Means algorithm clustering method, the data processing technique uses the KDD Knowledge Discovery in Database stage, the results obtained in this study are in 2019 the number of workers are still normal, in 2020 there will be a slight reduction in 2021, it will decrease by 3% from the last 3 years.

Keywords: *covid-19, K-means algorithm, statistical center*

PENDAHULUAN

Virus Covid-19 muncul di wuhan cina, pada desember 2019 tanggal 13 bulan januari 2020 terdapat kasus baru covid-19 diluar cina untuk pertama kalinya. Sedangkan di Indonesia, kasus pertama yang diumumkan presiden adalah pada tanggal 02 maret 2020. Selanjutnya, covid-19 dinyatakan sebagai pandemi oleh WHO pada tanggal 16 maret 2020. Sejak covid-19 dinyatakan sebagai pandemi, Indonesia menerapkan pembatasan social berskala besar sampai dengan mei 2020, termasuk di Sumatra selatan

terdapat beberapa wilayah yang menerapkan PSBB. Namun, setelah itu, perlahan-lahan kegiatan ekonomi dan social mulai dibuka kembali pada juni 2020. (Gunawan & Purwayoga, 2022)

Dengan adanya pandemic covid-19 tidak hanya masalah yang timbul, namun semua aspek dalam kehidupan ikut terdampak termasuk perekonomian. Perekonomian mulai menurun semenjak diberlakukannya pembatasan aktivitas. Hal ini terlihat dari pertumbuhan ekonomi yang menurun. Penurunan tersebut juga berdampak pada dinamika ketenaga kerjaan di Indonesia, termasuk di Sumatera Selatan. Tidak hanya pengangguran, penduduk usia kerja lainnya juga turut terdampak dengan adanya pandemic covid-19. (Studi et al., 2020)

Bagi Pemerintah Indonesia masalah yang terjadi adalah salah satu masalah yang sudah lama dan belum bisa diselesaikan. Masalah ini terjadi hampir di setiap wilayah salah satunya provinsi Sumatera Selatan. Ini disebabkan oleh kelangkaan kebutuhan dasar terhadap pendidikan dan pekerjaan. Pemerintah sendiri telah melakukan beberapa upaya dalam mengatasi covid-19 diantaranya melalui program PSBB. (Suhartini & Yuliani, 2021).

Masalah ini juga dipengaruhi oleh salah satu yaitu lapangan pekerjaan yang kurang memadai atau pekerjaan yang dimiliki mempunyai penghasilan yang tidak sesuai dengan tanggungan hidup seperti tanggungan dalam keluarga, anak dan lain sebagainya. Selain lapangan pekerjaan yang kurang memadai, ternyata salah satu penyebab masalah ini belum bisa diselesaikan adalah dari hasil pengamatan yang dilakukan bantuan yang diberikan pemerintah kepada penduduk tidak sesuai dengan apa yang mereka butuhkan.

Pada penelitian ini penulis ingin menerapkan Data Mining, Data mining adalah proses menemukan pola dan pengetahuan yang menarik dari suatu data dalam jumlah yang besar. Data mining menggambarkan sebuah pengumpulan teknik-teknik dengan tujuan untuk menemukan pola-pola yang tidak diketahui pada data yang telah dikumpulkan. penulis juga menerapkan Algoritma K-Means clustering. Dimana metode Algoritma K-Means Clustering ini bertujuan mengelompokkan data penduduk di provinsi Sumatera Selatan yang memang dikatankan tergolong penduduk yang terdampak covid-19. (Nabila et al., 2021)

Klasifikasi K-Means merupakan algoritma yang melakukan pengelompokkan data berdasarkan data yang terdekat dengan titik pusat *cluster*. (Alfianti, 2021) Data yang digunakan oleh penulis adalah data penduduk menurut jenis pekerjaan dari tahun 2019-2021 yang dimana data tersebut diperoleh dari Badan Pusat Statistik (BPS). dengan metode cluster ini diharapkan dapat memberikan informasi untuk pemerintah Sumatera Selatan.

METODE

Algoritma K-Means merupakan salah satu algoritma clustering yang sangat umum dalam mengelompokkan data sesuai dengan kesamaan karakteristik. Kelompok data yang dihasilkan disebut sebagai cluster/klaster. (Sembiring et al., 2022), Analisis cluster adalah metode yang dirancang untuk mengklasifikasikan objek berdasarkan karakteristik yang sama. (Widyadhana et al., 2021)

Metode yang digunakan dalam penelitian ini adalah metode deskriptif kualitatif, Metode deskriptif dapat diartikan sebagai prosedur pemecahan masalah yang diselidiki dengan menggambarkan keadaan subjek atau objek dalam penelitian dapat berupa orang, lembaga, masyarakat dan yang lainnya yang pada saat sekarang berdasarkan fakta-fakta yang tampak atau apa adanya. (Rizmayanti et al., 2021)

Metode pengumpulan data menggunakan metode sekunder, yaitu data-data yang akan menjadi sebuah bahan penelitian, yang diperoleh langsung dari tempat penelitian di Badan Pusat Statistik, yang dimana data-data tersebut yang akan menjadi sebuah bahan penelitian. (Matdoan et al., 2022)

1. Pengamatan (observasi)

Yaitu metode pengumpulan data dengan cara mengadakan tujuan secara langsung objek yang diteliti, Observasi dilakukan langsung pada Badan Pusat Statistik .

2. Studi Pustaka

Untuk mendapatkan data-data yang bersifat teoritis maka penulis melakukan pengumpulan data dengan cara membaca dan mempelajari buku-buku, jurnal penelitian, makalah ataupun referensi lain yang berhubungan dengan masalah yang dibahas.

Adapun untuk menganalisis data mining ini menggunakan knowledge discovery in database (KDD) yang terdiri dari beberapa tahapan, yaitu data selection, preprocessing, transformation, datamining dan evaluation.(Miskin, 2022)

1. Data Selection

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining akan disimpan dalam suatu berkas yang terpisah.

2. Pembersihan data

Sebelum proses data mining dapat dilaksanakan perlu dilakukannya proses pembersihan pada data yang menjadi focus KDD. Proses pembersihan mencakup antara lain membuang duplikasi data, memeriksa data yang tidak konsisten dan memperbaiki kesalahan pada data.

3. transformasi

Transformasi yang telah dipilih sehingga data tersebut sesuai untuk datamining. Proses ini merupakan proses yang sangat tergantung pada jenis atau pola yang dicari dalam basis data.

Sesudah pengumpulan data yang dilakukan, maka langkah selanjutnya menuju ketahap pengolahan data, data yang telah terkumpul dari Badan Pusat Statistik berdasarkan jenis pekerjaan dari tahun 2019 sampai dengan 2021. Berikut data pekerja yang berdasarkan jenis pekerjaan :

Table 1. Jenis pekerjaan tahun 2019-2021

Lapangan Pekerjaan Utama	2019	2020	2021
Pertanian	1.822	1.881	1.879
Pertambangan dan penggalian	64	62	61
Industri	247	241	264
Listrik, Gas dan Air	17	9	8
Konstruksi	210	202	222
Perdagangan	653	663	722
Angkutan Dan Komunikasi	188	179	160

Akomodasi Dan PMM	185	185	197
AKtivitas Dan Keuangan	38	36	30
Jasa perusahaan	35	43	41
ADM Pemerintah	131	141	145
Jasa Pendidikan	174	181	195
Jasa Kesehatan	55	63	61
Jasa Lainnya	145	166	160
Berusaha di bantu tidak tetap	885	900	903
Berusaha di bantu tetap	646	670	704
Buruh/Karyawan/Pegawai	113	91	99
Pekerjaan Bebas di pertanian	1.451	1.329	1.398
Pekerjaan Bebas di non pertanian	115	201	186
Pekerja Keluarga	108	142	138

Data diatas diatas merupakan data hasil seleksi yang telah dilakukan.

Transformasi data dilakukan untuk mengubah dan bertujuan agar data dapat diolah dengan menggunakan metode *K-Means Clustering*. Adapun variabel yang digunakan pada penelitian ini berupa data jenis pekerjaan berdasarkan tahun.

Table 2. Data Hasil Transfomasi

ID	JENIS PEKERJAAN	2019	2020	2021
K1	Pertanian	1.822	1.881	1.879
K2	Pertambangan dan penggalian	64	62	61
K3	Industri	247	241	264
K4	Listrik, Gas dan Air	17	9	8
K5	Kontruksi	210	202	222
K6	Perdagangan	653	663	722

K7	Angkutan Dan Komunikasi	188	179	160
K8	Akomodasi Dan PMM	185	185	197
K9	Aktivitas Dan Keuangan	38	36	30
K10	Jasa perusahaan	35	43	41
K11	ADM Pemerintah	131	141	145
K12	Jasa Pendidikan	174	181	195
K13	Jasa Kesehatan	55	63	61
K14	Jasa Lainnya	145	166	160
K15	Berusaha di bantu tidak tetap	885	900	903
K16	Berusaha di bantu tetap	646	670	704
K17	Buruh/Karyawan/Pegawai	113	91	99
K18	Pekerjaan Bebas di pertanian	1.451	1.329	1.398
K19	Pekerjaan Bebas di non pertanian	115	201	186
K20	Pekerja Keluarga	108	142	138

Data diatas merupakan hasil data yang telah di transformasikan dari exel ke rapidminer.

4. pengolahan data

pengolahan data jenis pekerja baru dilakukan setelah transformasi sehingga data baru bisa diolah menggunakan metode *k-means clustering*. Adapun langkah-langkah proses *algoritma k-means clustering* sebagai berikut:

1. dilakukan k dari jumlah clustering baru yang ingin dibentuk adapun cluster yang akan dibuat adalah tiga cluster.
2. menentukan titik pusat awal dari setiap cluster adapun penentuan titik pusat awal dalam penelitian ini ditentukan secara random dan titik pusat bisa dilihat pada table berikut ini:

Tabel 3. Titik pusat awal

Titik pusat	RANDOM		
CENTROID 1	17	61	8
CENTROID 2	115	99	145
CENTROID 3	903	30	41

3. Menghitung jarak data ke pusat cluster antara objek ke centroid terdekat akan menjadi cluster yang ikuti oleh data tersebut. Perhitungan jarak *euclidian* dapat dilakukan dengan persamaan berikut:

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}$$

persamaan diatas digunakan karena atribut yang digunakan berjumlah 3 atribut sebagai contoh akan dihitung jarak dari data pekerja tahun 2019 ke pusat pertama ke pusat cluster pertama dengan persamaan.

$$d(1,1) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}$$

$$d(1,1) = \sqrt{(1822 - 17)^2 + (1881 - 61)^2 + (1879 - 8)^2}$$

????????

Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data pekerja tahun 2019 ke pusat cluster pertama adalah 317,5. Jarak data ketenagakerjaan ke pusat cluster kedua dengan persamaan:

$$d(1,2) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}$$

$$d(1,2) = \sqrt{(1822 - 115)^2 + (1881 - 99)^2 + (1879 - 145)^2}$$

????????

Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data pekerja tahun 2020 ke pusat *cluster* pertama adalah 3016,0. Jarak data jumlah ketenaga kerjaan ke pusat *cluster* ketiga dengan persamaan:

$$d(1,3) = \sqrt{(p1 - q1)^2 + (p2 - q2)^2 + (p3 - q3)^2}$$

$$d(1,3) = \sqrt{(1822 - 903)^2 + (1881 - 30)^2 + (1879 - 41)^2}$$

$$2765,7$$

Dari hasil perhitungan diatas didapatkan hasil bahwa jumlah pekerja tahun 2021 ke pusat *cluster* ketiga adalah 2765,7.

Berdasarkan hasil perhitungan diatas dapat disimpulkan bahwa jarak data pekerja yang paling mendekat adalah cluster 3, sehingga data k1 dimasukkan kedalam cluster 3 hasil perhitungan selengkap nya untuk 20 sample data pekerja dapat dilihat pada table berikut ini.

Tabel 4. Hasil perhitungan setiap data ke setiap cluster iterasi 1

ID	JARAK KE CENTROID			JARAK TERDEKAT	CLUSTER DIIKUTI
	1	2	3		
K1	3173,5	3016,0	2765,7	2765,7	3
K1	70,8	105,0	839,8	70,8	1
K3	388,4	227,5	724,3	227,5	2
K4	52,0	191,0	886,9	52,0	1
K5	320,8	159,9	736,6	159,9	2
K6	1129,9	969,8	962,8	962,8	3
K7	257,4	109,3	740,0	109,3	2
K8	281,6	122,5	750,9	122,5	2
K9	39,4	152,1	865,1	39,4	1
K10	41,7	142,7	868,1	41,7	1
K11	195,4	44,9	786,8	44,9	2
K12	272,1	112,7	760,2	112,7	2
K13	65,2	109,3	848,9	65,2	1
K14	224,8	74,9	779,2	74,9	2

K15	1502,8	1345,0	1224,9	1224,9	3
K16	1118,5	959,4	956,7	956,7	3
K17	135,6	46,7	794,5	46,7	2
K18	2365,6	2206,3	1956,8	1956,8	3
K19	246,8	109,9	819,3	109,9	2
K20	178,2	44,1	808,7	44,1	2

1. Setelah semua data ditempatkan dalam cluster yang terdekat, kemudian hitung kembali pusat cluster yang baru berdasarkan rata-rata anggota yang ada pada cluster tersebut.
2. Jika titik centroid yang baru konvergen dengan centroid yang lama maka hentikan iterasi. Jika tidak iterasi dilanjutkan ke yang berikutnya.

Selanjutnya mengelompokan hasil cluster pada iterasi pertama yang belum konvergen. Untuk membangkitkan centroid yang baru dengan rumus berikut:

$$C = \frac{\sum x_m}{n}$$

Dimana :

C : centroid data

m : Anggota Data yang termasuk dalam centroid tertentu

n : jumlah data yang menjadi anggota centroid tertentu

pada pengelompokan data dengan menggunakan rumus diatas didapatkan titik pusat cluster dengan nilai sebagai berikut:

Contoh perhitungan data sampel 15 orang mahasiswa pada Cluster 1 adalah berikut:

$$\frac{64 + 17 + 38 + 55}{4} = 43,5$$

$$\frac{62 + 9 + 36 + 63}{4} = 42,5$$

$$\frac{61 + 8 + 30 + 61}{4} = 40$$

Demikian perhitungan pada cluster 2 dan cluster 3. Pada pengelompokan data dengan menggunakan rumus diatas didapatkan titik pusat cluster dengan nilai sebagai berikut:

tabel 5. Titik pusat iterasi 1 setelah cluster

titik pusat 2	BARU		
c0	43,5	42,5	40
c1	150,091	161,0909	164,2727
c2	1091,4	1088,6	1121,2

Karena centroid baru yang digunakan belum konvergen maka iterasi harus dilanjutkan dalam penelitian ini, iterasi terjadi sebanyak 2 kali iterasi karena kondidi cluster sudah mencapai konvergen dan proses iterasi pun berhenti adapun hasil akhir clustering dari 20 data pekerja dapat dilihat pada tabel berikut :

Tabel 6. Hasil perhitungan data kesetiap cluster pada iterasi-2

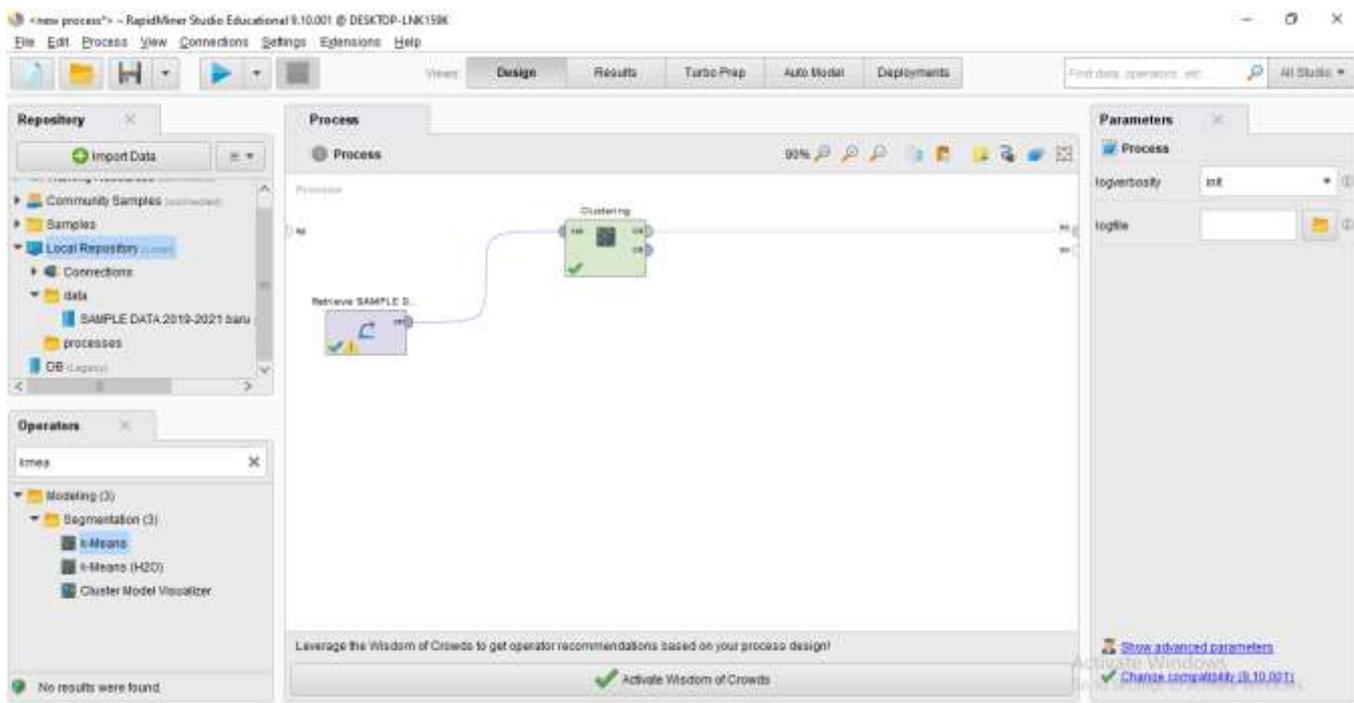
ID	JARAK KE CENTROID			JARAK TERDEKAT	CLUSTER DIIKUTI
	1c	2c	3c		
1	3150,4	2948,5	1317,5	1317,5	3
2	35,2	167,0	1798,2	35,2	1
3	361,9	160,4	1471,8	160,4	2
4	53,4	255,5	1886,6	53,4	1
5	293,7	92,7	1540,0	92,7	2
6	1105,3	903,3	729,9	729,9	3
7	232,2	42,1	1602,3	42,1	2
8	254,9	53,5	1578,7	53,5	2
9	13,1	215,0	1846,2	13,1	1
10	8,6	205,9	1837,4	8,6	1
11	168,5	33,8	1665,3	33,8	2
12	245,4	43,7	1588,5	43,7	2
13	31,5	171,3	1802,8	31,5	1
14	199,9	8,3	1634,2	8,3	2
15	1479,3	1277,4	354,7	354,7	3
16	1094,4	892,3	740,0	740,0	3
17	103,3	102,7	1731,3	102,7	2
18	2341,0	2139,7	513,5	513,5	3
19	227,0	57,4	1617,3	57,4	2
20	153,8	53,2	1682,2	53,2	2

Pada iterasi ke 2, titik pusat dari setiap cluster sudah tidak berubah dan tidak ada lagi data yang berpindah dari satu cluster ke cluster yang lain.

4. Hasil dan pembahasan

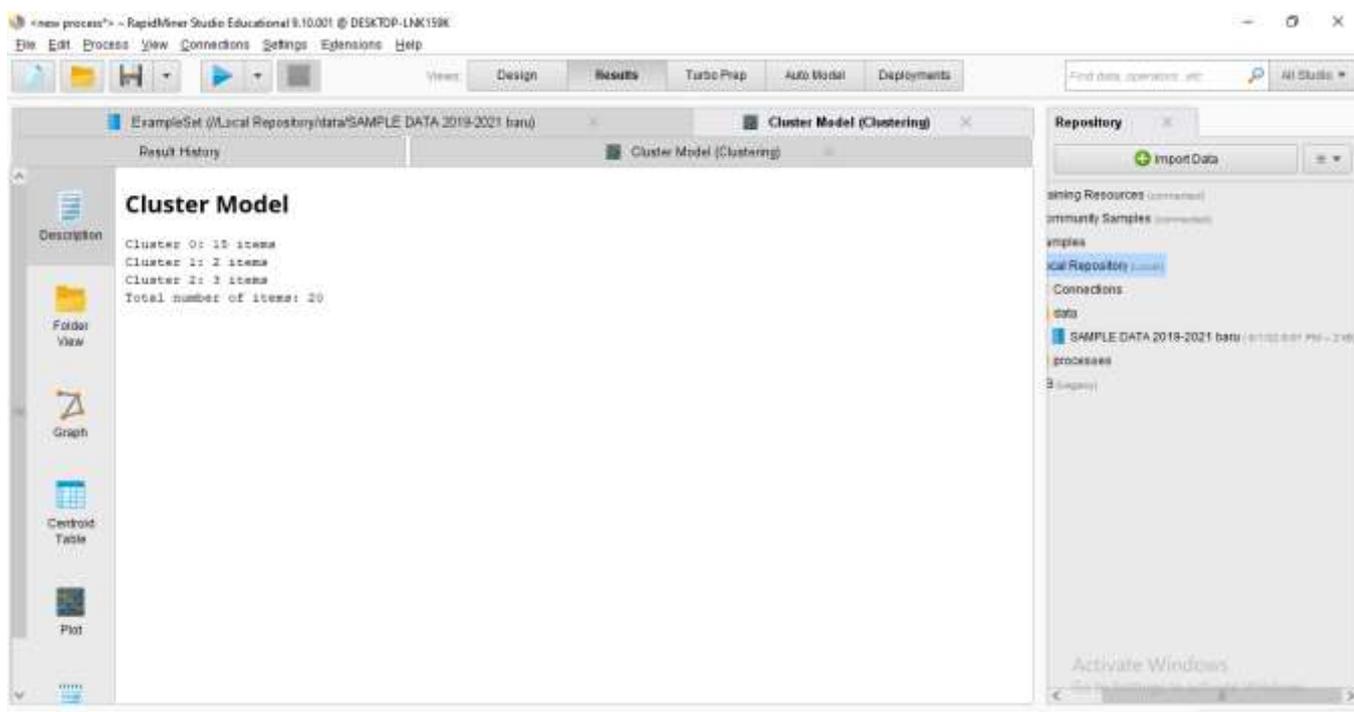
4.1 hasil iterasi dengan metode k-means

adapun pengolahan data pekerja dengan menggunakan algoritma k-means clustering dan software rapid-miner dapat dilihat pada gambar berikut ini:



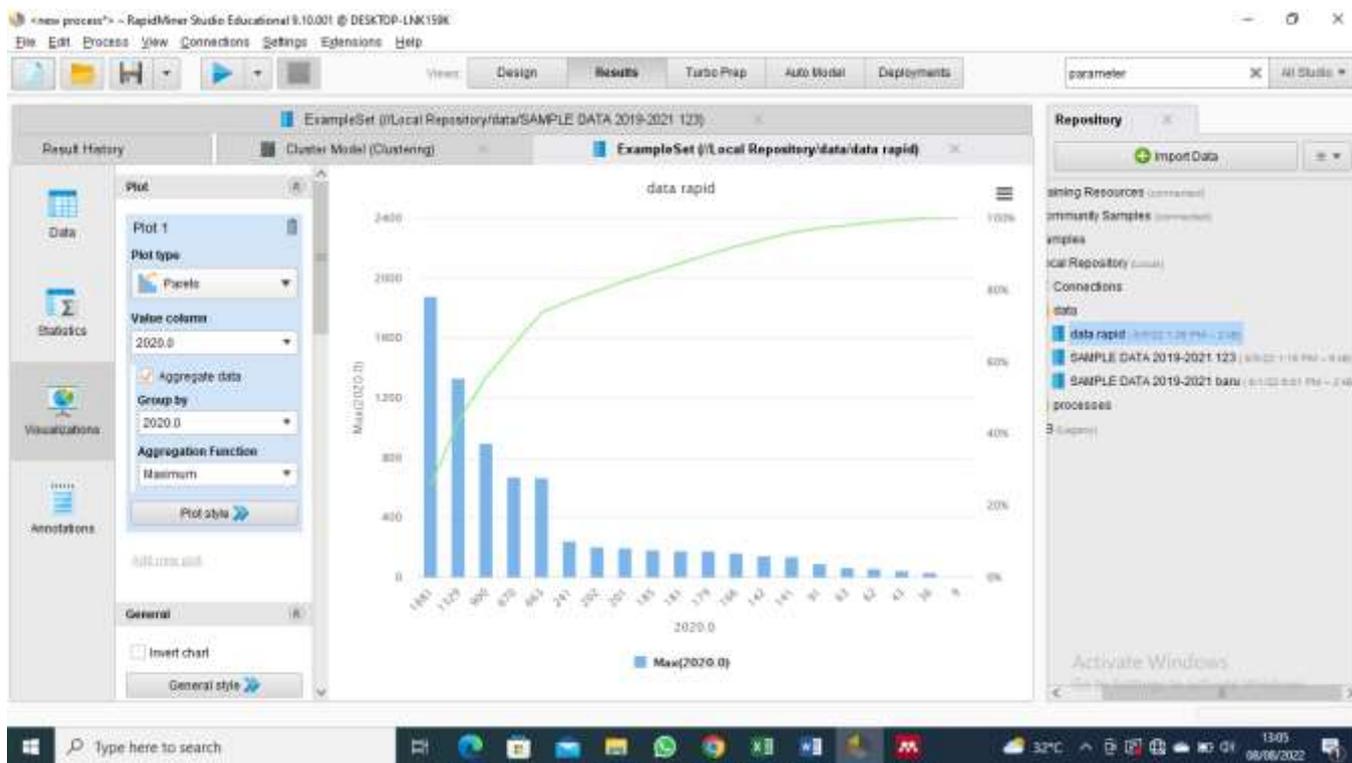
Gambar 1. Proses penerapan algoritma K-Means

Gambar di atas adalah proses penerapan algoritma K-Means



Gambar 2. Cluster model

Gambar di atas adalah hasil dari algoritma K-Means dengan menggunakan perhitungan tools RapidMiner di dapatkan cluster₀ = 15 item, cluster₁ = 2 item dan cluster₃ = 3 item jadi totalnya ada 20 item.



Gambar 3. Hasil K-Means clustering pada rapidminer

SIMPULAN

Dari hasil yang didapatkan, cluster pekerjaan di tahun 2019-2021 dapat disimpulkan melalui perhitungan menggunakan metode K-Means, peningkatan dan penurunan yang terjadi dari tahun ketahun dapat dilihat dari analisis yang telah dilakukan, terjadi penurunan pada tahun 2020 karena banyak pekerja yang terkena dampak dari covid-19. Sehingga tingkat lapangan pekerjaan mengalami penurunan pada tahun 2020, sejak tahun 2021 agustus mulai ada sedikit peningkatan walau tidak terlalu banyak kurang lebih sekita 3% dari jumlah sebelumnya.

DAFTAR PUSTAKA

- Alfianti, Z. I. (2021). Pengelompokan Wilayah Penyebaran Covid-19 Di Kabupaten Karawang Menggunakan Algoritma K-Means. *Jurnal Ilmiah Informatika Komputer*, 26(2), 111–122. <https://doi.org/10.35760/ik.2021.v26i2.4155>
- Gunawan, H., & Purwayoga, V. (2022). Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Potensi Penyebaran Virus Corona Di Kota Cirebon. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 11(1), 1–8. <https://doi.org/10.32736/sisfokom.v11i1.1316>
- Matdoan, Y., Matdoan, U. A., & Far-far, M. S. (2022). *Algoritma K-Means Untuk Klasifikasi Provinsi di Indonesia Berdasarkan Paket Pelayanan Stunting Pendahuluan*. 1(2), 41–46.
- Miskin, P. K. (2022). *IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING DALAM ALGORITMA K-MEANS CLUSTERING DALAM*.
- Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/JTSI>
- Rizmayanti, A. I., Hidayati, N., Nugraha, F. S., & Gata, W. (2021). Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decision Tree (Studi Kasus Smk Multicomp Depok). *Swabumi*, 9(1), 9–18. <https://doi.org/10.31294/swabumi.v9i1.8363>
- Sembiring, C. S. D. B., Hanum, L., & Tamba, S. P. (2022). Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus Ftik Unpri). *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 5(2), 80–85.

<https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2393>

Studi, P., Informasi, S., & Buana, U. M. (2020). *Program Studi Sistem Informasi Universitas Mercu Buana Yogyakarta 2020* (Issue 01).

Suhartini, S., & Yuliani, R. (2021). Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur. *Infotek : Jurnal Informatika Dan Teknologi*, 4(1), 39–50. <https://doi.org/10.29408/jit.v4i1.2986>

Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). Perbandingan Analisis Kluster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 584–594. <https://journal.unnes.ac.id/sju/index.php/prisma/>