



Penerapan Algoritma K-Means Dalam Pengelompokan Kepadatan Penduduk Menurut Kecamatan di Kabupaten Simalungun

Devi Gultom^{1,*}, Indra Gunawan¹, Ika Purnamasari¹, Sundari Retno Andani², Zulia Almaida Siregar²

¹ STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

² AMIK Tunas Bangsa, Pematangsiantar, Indonesia

Email: ¹devigultom30@gmail.com, ²indra@amiktunasbangsa.ac.id, ³ikapurnamasari1319@gmail.com,

⁴sundari.ra@amiktunasbangsa.ac.id, ⁵zuliaalmaida@yahoo.com

Email Penulis Korespondensi: devigultom30@gmail.com

Abstrak—Salah satu persoalan yang terkait dengan kependudukan yang masih harus dihadapi oleh Simalungun yaitu masalah ketidakseimbangan pembagian penduduk. Pembagian penduduk yang tidak menyeluruh mendatangkan masalah pada kepadatan penduduk dan tekanan penduduk di suatu wilayah. Penelitian ini menggunakan sumber data dari Badan Pusat Statistik (BPS) Simalungun. Data yang digunakan dalam penelitian ini yakni data dari tahun 2015-2019 yang terdiri dari 32 Kecamatan. Oleh karena itu, peneliti melakukan pemanfaatan algoritma K-Means dalam clustering 32 Kecamatan di Kabupaten Simalungun. Data akan diolah dengan melakukan clustering dalam 3 cluster yaitu cluster tingkat wilayah penduduk yang tinggi, cluster tingkat wilayah penduduk yang sedang dan cluster tingkat wilayah penduduk yang rendah. Penelitian ini mengelompokkan kepadatan penduduk dengan menggunakan software Ms. Excel dan RapidMiner. Proses iterasi berlangsung sebanyak 5 kali sehingga hasil yang diperoleh 4 Kecamatan dengan cluster tingkat wilayah penduduk yang tinggi (C1), 11 Kecamatan cluster tingkat wilayah penduduk yang sedang (C2) dan 17 Kecamatan cluster tingkat wilayah penduduk yang rendah (C3).

Kata Kunci: Penduduk Simalungun; Data Mining; Algoritma K-Means; Clustering; RapidMiner

Abstract—One of the problems related to population that still has to be faced by Simalungun is the problem of the imbalance in the distribution of the population. Incomplete division of the population brings problems to population density and population pressure in an area. This study uses data sources from the Central Statistics Agency (BPS) Simalungun. The data used in this study is data from 2015-2019 which consists of 32 Districts. Therefore, the researchers used the K-Means algorithm in clustering 32 sub-districts in Simalungun Regency. The data will be processed by clustering in 3 clusters, namely clusters with high population levels, clusters with moderate population levels and clusters with low population levels. The iteration process takes 5 times so that the results obtained are 4 sub-districts with high population level clusters (C1), 11 cluster sub-districts with moderate population level (C2) and 17 cluster sub-districts with low population level (C3).

Keywords: Simalungun Population; Data Mining; K-Means Algorithm; Clustering; RapidMiner

1. PENDAHULUAN

Kepadatan penduduk adalah jumlah rata-rata penduduk pada setiap wilayah satu kilometer persegi. Jumlah kepadatan penduduk tiap-tiap wilayah tidak sama. Kepadatan penduduk dipengaruhi oleh dua faktor antara lain faktor *intern* daerah seperti pertumbuhan penduduk yang relatif tinggi (kelahiran lebih tinggi dibandingkan kematian) dan wilayah yang strategis maupun faktor *eksternal* seperti banyaknya penduduk berpindah dan wilayah di sekitarnya yang kurang produktif. Salah satu persoalan yang terkait dengan kependudukan yang masih harus dihadapi oleh Simalungun yaitu masalah ketimpangan pembagian penduduk. Pembagian wilayah yang tidak seimbang menimbulkan masalah pada kepadatan penduduk dan himpitan penduduk di suatu wilayah [1].

Tiap wilayah harus mampu menjamin penduduknya agar dapat mengakses layanan yang terkait dengan kebutuhan pokok secara merata karena pembangunan sumber daya manusia sangat penting bagi pembangunan saat ini [2]. Pengolahan Data mining salah satu dapat menyelesaikan masalah pada penulis tersebut. Data mining merupakan sebuah pusat dari proses KDD dimana membentuk model dan menciptakan pola yang belum diketahui. Data Mining dapat dialjabarkan sebagai pengorganisasian proses untuk proses data yang besar [3].

Data mining menggambarkan penyelesaian masalah dengan menganalisa data yang ada pada database, dimana data tersimpan secara elektronik dan pencariannya dilakukan manual seperti pada komputer. Salah satu cara dapat menggambarkan karakteristik sekelompok wilayah berdasarkan satu atau lebih variabel antara lain teknik *clustering*. *K-Means* merupakan salah satu metode clustering/ pengelompokan non hirarki. Teknik pengelompokan datanya sederhana dan cepat pada data yang berukuran besar [4].

Metode ini bertujuan membagi n observasi ke dalam K cluster dimana tiap observasi menjadi anggota dari kluster dengan nilai rata-rata terdekat, melalui proses perulangan (iterasi) hingga pengelompokan tersebut konvergen. Penelitian yang dilakukan [5] oleh mengatakan Kepadatan penduduk ini dapat membawa pengaruh buruk bagi masyarakat jika tidak segera ditangani, salah satunya adalah meningkatnya kasus tindakan kriminal. Pada penelitian terdahulu, [6] Algoritma *K-Means* merupakan algoritma pengelompokkan sejumlah data menjadi kelompok-kelompok data tertentu. Pada penelitian ini data penjualan dikelompokkan menjadi 3 yaitu data penjualan rendah, data penjualan sedang dan data penjualan tinggi.

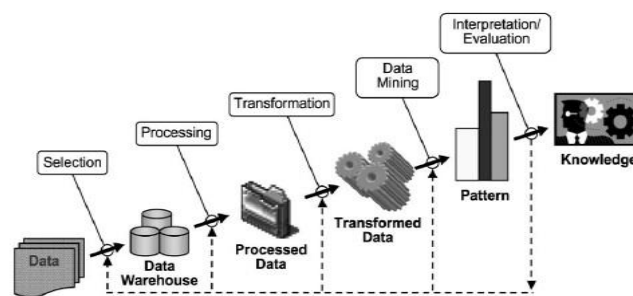
2. METODOLOGI PENELITIAN

2.1 Data Mining

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Perlu diingat bahwa kata “mining” sendiri berarti usaha untuk mendapatkan sedikit data berharga dari sejumlah besar data dasar [7]. Teknik penambangan data digunakan untuk memeriksa basis data berukuran besar sebagai aturan untuk menemukan model yang baru dan bermanfaat. Data Mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan Data Mining adalah kenyataan bahwa Data Mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu [8]. Bermula tentang disiplin ilmu, Data Mining bertujuan untuk membenarkan teknik tradisional sehingga bisa menyelenggarakan:

- Jumlah data yang sangat besar
- Dimensi data yang tinggi
- Data yang heterogen dan berbeda sifat

Berikut ini tahapan yang dilakukan pada proses data mining yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik. Secara detail dijelaskan sebagai berikut :



Gambar 1. Tahapan Proses Data Mining

2.2 Clustering

Clustering mengacu pada pengelompokan seperti catatan, pemeriksaan, atau peninjauan dan membentuk kelas objek-objek yang memiliki kesamaan. *Cluster* adalah kumpulan dari pemeriksaan yang memiliki kesamaan satu sama lain, dan berbeda dengan pemeriksaan di kluster lain [9]. *Clustering* memandu agar membagi seluruh kumpulan data menjadi kelompok-kelompok yang relatif memiliki kesamaan, di mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kesamaan *record* dalam kelompok lain akan bernilai minimal.

2.3 Algoritma K-Means

K-Means mengadakan suatu algoritma yang digunakan dalam bagian pengelompokkan secara pertisi yang memisahkan data ke dalam kelompok yang berbeda-beda. Algoritma *K-Means* bisa menurunkan jarak antara data ke *clusternya* [10]. Metode *K-Means* yakni metode clustering yang membagi data kedalam satu atau lebih kluster, data yang karakteristik sama dikelompokkan satu kluster yang sama dan karakteristik yang berbeda dikelompokkan kedalam kelompok yang lain [11]. Algoritma *K-means* akan memeriksa setiap komponen dalam anggota data tersebut dan menandai anggota tersebut ke dalam salah satu utama kluster yang telah dideskripsikan sebelumnya tergantung dari jarak minimum antar anggota dengan tiap-tiap pusat kluster. Posisi kluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap-tiap *cluster* dan terakhir akan terbentuk kluster baru [12].

K-means yakni metode analisis kelompok yang mengarah pada pembagian kelompok dan setiap objek pengamatan dimiliki oleh suatu kelompok dengan rata-rata terdekat, sehingga menghasilkan keanggotaan yang berbeda pula. [13] Penggolongan data dengan metode *K-Means* dilakukan dengan algoritma, yakni :

- Tentukan banyak kluster k .
- Berikan nilai awal pusat *cluster* sebanyak k , umunya secara acak dari data yang tersedia.
- Masukkan setiap data ke dalam kluster yang memiliki *centroid* terdekat. Gunakan *Euclidean distance* untuk menghitung jarak dari setiap data ke setiap *centroid* dari kluster dengan rumus berikut ini:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \quad (1)$$

Keterangan:

d_{ik} = jarak antara data ke- i ke *centroid cluster* ke- k

m = jumlah atribut

x_j = data ke- j

c_k = *centroid cluster* ke- k

- Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan sebagai berikut,



$$v = \sum_{i=1}^n \frac{x_i}{n} \tag{2}$$

Keterangan:

v = centroid dari cluster

xi = data ke-i. i = 1,2,3...n

n = banyaknya jumlah data yang menjadi anggota cluster

e) Apabila, nilai centroid ada perubahan maka kembali ke langkah 3

2.4 RapidMiner 5.3

RapidMiner suatu perangkat lunak yang bersifat terbuka (open source). RapidMiner yakni sebuah solusi untuk melakukan penjabaran terhadap data mining, text mining dan uraian prediksi. RapidMiner yakni platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan yang bernama sama dengan pihak yang memberi lingkungan terintegrasi untuk persiapan data, penggalian teks dan analisis prediktif. RapidMiner dilengkapi dengan satu prosesor logika dan 10.000 baris data [14]. RapidMiner adalah sebuah tools yang digunakan dalam teknik yang berada di lingkungan machine learning, data mining, text mining dan predictive analytics [15].

3. HASIL DAN PEMBAHASAN

3.1 Perhitungan Dengan Algoritma K-Means

Pada penelitian ini data yang digunakan adalah kepadatan penduduk di Kabupaten Simalungun menurut Kecamatan yang diambil dari tahun 2015-2019 yang diperoleh dari Badan Pusat Statistik (BPS) Simalungun. Dalam melakukan clustering, data yang di terima akan dihitung terlebih dahulu dengan menggunakan algoritma K-Means.

Tabel 1. Data Kepadatan Penduduk

No	Kecamatan	Kepadatan Penduduk				
		2015	2016	2017	2018	2019
1	Siantar	890	896	902	908	913
2	Pematang Silimahuta	135	136	137	138	138
3	Purba	137	139	141	142	144
4	Haranggaol Horison	124	124	124	124	125
5	Dolok Pardamean	157	157	190	191	191
6	Sidamanik	341	342	343	344	345
7	Pamatang Sidamanik	121	121	121	122	122
8	Girsang Sipangan Bolon	114	115	115	116	116
9	Bandar	677	685	692	699	705
10	Hatonduhan	63	64	64	64	64
11	Dolok Panribuan	123	124	124	124	124
12	Jorlang Hataran	167	167	168	168	168
13	Panei	283	285	286	287	288
14	Panombean Panei	264	265	266	266	267
15	Raya	97	97	99	100	100
16	Dolog Masagal	0	0	92	93	93
17	Dolok Silau	47	48	48	48	48
18	Silau Kahean	76	77	77	77	78
19	Raya Kahean	87	87	87	88	88
20	Tapian Dolok	340	343	347	350	353
21	Dolok Batu Nanggar	379	380	382	383	385
22	Silimakuta	208	213	217	221	225
23	Gunung Malela	356	359	362	365	367
24	Gunung Maligas	539	543	548	552	556
25	Hutabayu Raja	155	156	156	157	157
26	Jawa Maraja Bah Jambi	557	565	572	579	586
27	Pematang Bandar	359	359	360	360	360
28	Bandar Huluan	246	247	247	248	249
29	Tanah Jawa	273	273	274	275	275
30	Bandar Masilam	272	273	273	274	275
31	Bosar Maligas	141	142	143	144	144
32	Ujung Padang	180	181	181	181	182

a) Pada tahap ini didapatkan hasil :

Jumlah Cluster : 3

Jumlah Data : 32

b) Menentukan nilai centroid secara random

Tabel 2. Centroid Data Awal

	2015	2016	2017	2018	2019
C1	890	896	902	908	913
C2	273	273	274	275	275
C3	0	0	92	93	93

c) Menghitung *centroid* terdekat :

Kecamatan Siantar :

$$= \sqrt{(890-890)^2 + (896-896)^2 + (902-902)^2 + (908-908)^2 + (913-913)^2}$$

$$D_{(1,1)} = 0$$

$$= \sqrt{(890-273)^2 + (896-273)^2 + (902-274)^2 + (908-275)^2 + (913-275)^2}$$

$$D_{(1,2)} = 1403,90$$

$$D_{(1,3)} = \sqrt{(890-0)^2 + (896-0)^2 + (902-92)^2 + (908-93)^2 + (913-93)^2}$$

$$= 1894,11$$

Kecamatan Ujung Padang :

$$D_{(1,1)} = \sqrt{(180-890)^2 + (181-896)^2 + (181-902)^2 + (181-908)^2 + (182-913)^2}$$

$$= 1611,85$$

$$D_{(1,2)} = \sqrt{(180-273)^2 + (181-273)^2 + (181-274)^2 + (181-275)^2 + (182-275)^2}$$

$$= 207,96$$

$$D_{(1,3)} = \sqrt{(180-0)^2 + (181-0)^2 + (181-92)^2 + (181-93)^2 + (182-93)^2}$$

$$= 297,90$$

Hasil dari keseluruhan perhitungan mampu di amati hasil pengelompokan menurut iterasi 1, sebagai berikut :

Tabel 3. Hasil Perhitungan Jarak Pusat Cluster Iterasi 1

No	Kecamatan	C1	C2	C3	Jarak Terpendek
1	Siantar	0,00	1403,90	1894,11	0,00
2	Pematang Silimahuta	1710,67	306,79	206,87	206,87
3	Purba	1702,14	298,31	213,29	213,29
4	Haranggaol Horison	1738,86	334,97	183,74	183,74
5	Dolok Pardamean	1620,41	219,27	279,48	219,48
6	Sidamanik	1249,61	154,30	650,19	154,30
7	Pamatang Sidamanik	1745,11	341,23	178,34	178,34
8	Girsang Sipangan Bolon	1758,97	355,09	166,76	166,76
9	Bandar	470,04	934,00	1424,55	470,04
10	Hatonduhan	1873,91	470,02	102,62	102,62
11	Dolok Panribuan	1739,75	335,86	182,90	182,90
12	Jorlang Hataran	1641,81	237,92	269,82	237,92
13	Panei	1377,49	26,48	524,03	26,48
14	Panombean Panei	1422,68	18,81	480,00	18,81
15	Raya	1796,08	392,21	137,71	137,71
16	Dolog Masagal	1894,11	498,43	0,00	0,00
17	Dolok Silau	1909,68	505,80	102,46	102,46
18	Silau Kahean	1844,39	440,51	111,40	111,40
19	Raya Kahean	1821,14	417,25	123,34	123,34
20	Tapian Dolok	1241,49	162,56	657,21	162,56
21	Dolok Batu Nanggar	1162,84	241,07	735,87	241,07
22	Silimakuta	1531,72	128,41	371,57	128,41
23	Gunung Malela	1207,51	196,45	691,07	196,45
24	Gunung Maligas	792,03	611,90	1103,79	611,90
25	Hutabayu Raja	1667,30	263,41	246,27	246,27
26	Jawa Maraja Bah Jambi	737,92	666,23	1157,20	666,23
27	Pematang Bandar	1212,52	191,41	687,14	191,41
28	Bandar Huluan	1463,37	59,49	440,35	59,49

No	Kecamatan	C1	C2	C3	Jarak Terpendek
29	Tanah Jawa	1403,90	0,00	498,43	0,00
30	Bandar Masilam	1405,24	1,73	497,15	1,73
31	Bosar Maligas	1697,25	293,37	218,74	218,74
32	Ujung Padang	1611,85	207,96	297,90	207,96

d) Menghitung *Centroid* Baru

Menentukan *centroid* baru dengan menghitung jumlah yang terpilih pada *cluster* kemudian membagikannya sebanyak jumlah *cluster* yang terpilih. Demikian di peroleh dari hasil jarak dari setiap objek pada iterasi ke-1 maka dilanjutkan ke iterasi ke-2 menurut perhitungan di bawah ini.

Tabel 4. Centroid Baru

	2015	2016	2017	2018	2019
C1	783,5	790,5	797	803,5	809
C2	307,5	309,5	313,5	315,25	317
C3	101,4	102,1	109,1	109,7	110,0

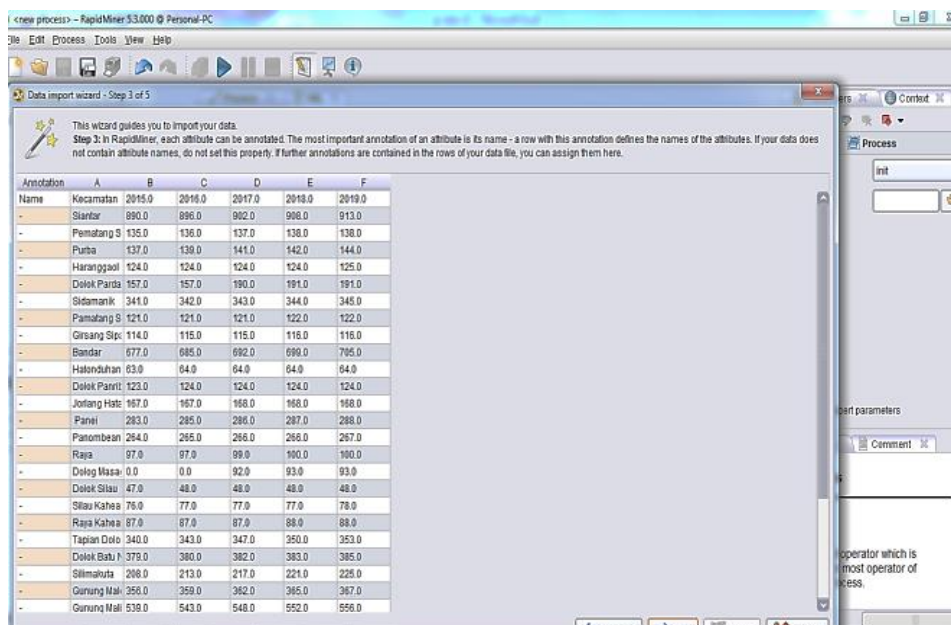
Perhitungan akan berhenti apabila tahapan iterasi sudah mendapatkan hasil yang sama, maka proses algoritma *K-Means* berhenti. Pada penelitian ini, proses iterasi yang dilakukan sampai pada iterasi ke 5. Hasil akhir clustering iterasi ke 5 dapat di liat pada tabel berikut:

Tabel 5. Hasil Cluster Iterasi 5

Cluster	Kecamatan	Jumlah
C1	Siantar, Bandar, Gunung Maligas dan Jawa Maraja Bah Jambi.	4
C2	Sidamanik, Panei, Panombeian Panei, Tapian Dolok, Dolok Batu Nanggar, Silimakuta, Gunung Malela, Pematang Bandar, Bandar Hulan, Tanah Jawa dan Bandar Masilam.	11
C3	Pematang Silimahuta, Purba, Haranggaol Horison, Dolok Pardamean, Pematang Sidamanik, Girsang Sipangan Bolon, Hatonduhan, Dolok Panribuan, Jorlang Hataran, Raya, Dolog Masagal, Dolok Silau, Silau Kahean, Raya Kahean, Hutabayu Raja, Bosar Maligas dan Ujung Padang.	17

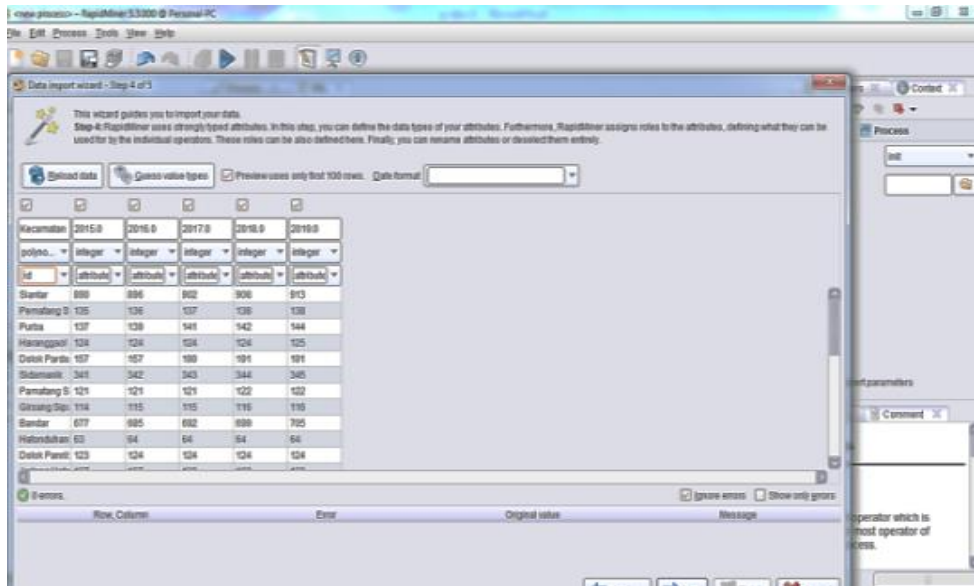
3.2 Uji Coba Software RapidMiner 5.3

Halaman *input* data merupakan halaman yang berfungsi untuk memasukan data baru ke dalam database untuk selanjutnya di proses lebih lanjut oleh aplikasi, terlebih dahulu *importing* data dari yang sudah di transformasi terlebih dahulu ke dalam *Microsoft Excel* agar dapat di proses menggunakan *RapidMiner*. Klik pada bagian kiri bawah tab *repositories*, lalu pilih “*Import Excel Sheet*”. Pada data *import wizard* kemudian pilih tempat data yang akan digunakan untuk di *import*. Kemudian klik *next*. Pilih *sheet* pada data serta blok data yang akan di proses. Pilih *next* dan akan tampil seperti berikut:



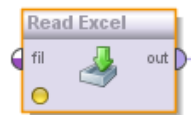
Gambar 2. Data Import Tahap 3

Pada form data *import*, ubah salah satu atribut menjadi Id. Pada penelitian ini atribut yang diubah menjadi Id adalah kolom Kecamatan.



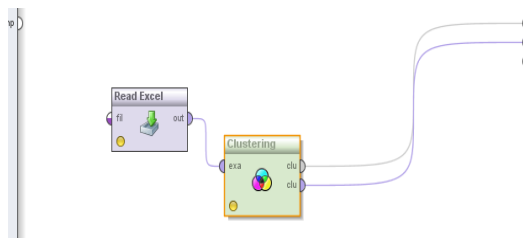
Gambar 3. Data Import Tahap 4

Setelah ditentukan *Id* yang diubah maka klik *Finish*, selanjutnya akan kembali ke halaman *process* awal dengan pertanda dari *operator read excel* berwarna kuning yang sebelumnya berwarna merah seperti gambar dibawah ini :



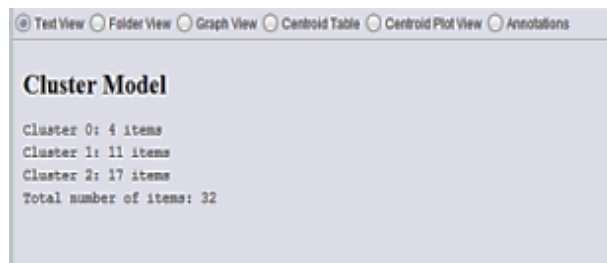
Gambar 4. Tampilan Read Excel yang Berhasil di Import

Langkah berikutnya *drag* dan *drop* operator *K-Means* dengan menentukan nilai *cluster* ($k=3$) kemudian hubungkan pada *output*.



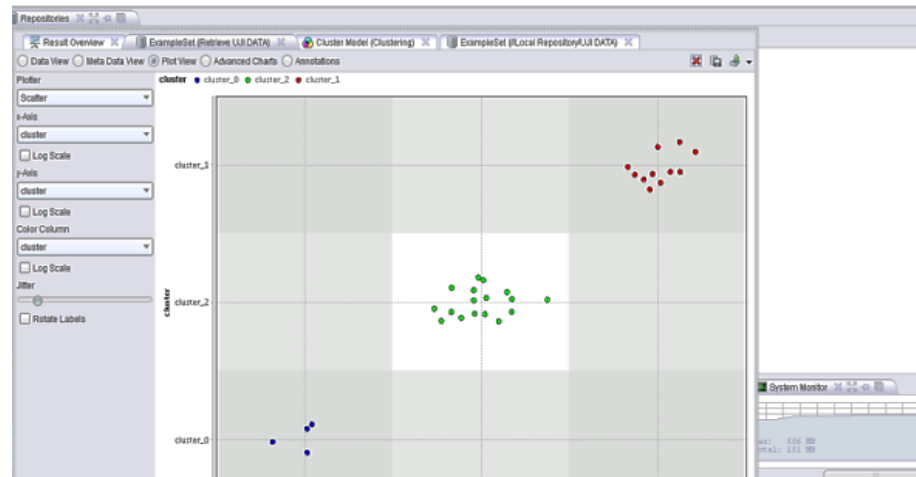
Gambar 5. Menghubungkan Read Excel yang Berhasil dengan Operator K-Means

Pada gambar 5 yakni proses menghubungkan *read excel* dengan operator *K-Means* dan *output* yang akan di eksekusi. Berikutnya, klik *tool run* yang berbentuk segitiga kesamping berwarna biru. Pada saat *tool run* di klik setelah keluar hasil *cluster* seperti gambar di bawah ini :



Gambar 6. Data Hasil Clustering Menggunakan K-Means

Berdasarkan pada gambar 6 menjelaskan bahwa terdapat 3 *cluster* yang dimulai dari *cluster 0*, *cluster 1* dan *cluster 2*. Dimana *cluster 0* merupakan *cluster* tinggi, *cluster 1* merupakan *cluster* sedang dan *cluster 2* merupakan *cluster* rendah.



Gambar 7. Grafik Clustering

4. KESIMPULAN

Bedasarkan pembahasan diatas dapat disimpulkan bahwa perhitungan secara manual algoritma *K-Means clustering* dengan menggunakan *RapidMiner* memiliki hasil yang sama. Data yang diolah menggunakan *Microsoft Excel* untuk memperoleh kepadatan penduduk di Kabupaten Simalungun menerapkan 3 *clustering K-Means* dimana nilai *centroid* yang berhenti pada iterasi ke-5 dalam 3 *cluster* yaitu *cluster* tinggi, sedang dan rendah. Hasil yang didapatkan dari algoritma *K-Means clustering* yang di proses ke dalam *RapidMiner* menyimpan nilai yang sama yaitu menghasilkan 3 *cluster* yaitu *cluster* tinggi (C1) sebanyak 4 Kecamatan yaitu Siantar, Bandar, Panung Maligas dan Jawa Maraja Bah Jambi. *Cluster* sedang (C2) sebanyak 11 Kecamatan yaitu Sidamanik, Panei, Panombean Panei, Tapan Dolok, Dolok Batu Nanggar, Silimakuta, Gunung Malela, Pematang Bandar, Bandar Hulan, Tanah Jawa dan Bandar Masilam. *Cluster* rendah (C3) sebanyak 17 Kecamatan yaitu Pematang Silimahuta, Purba, Haranggaol Horison, Dolok Pardamean, Pematang Sidamanik, Girsang Sipangan Bolon, Hatonduhan, Dolok Panribuan, Jorlang Hataran, Raya, Dolog Masagal, Dolok Silau, Silau Kahean, Raya Kahean, Hutabayu Raja, Bosar Maligas dan Ujung Padang.

REFERENCES

- [1] BPS Kabupaten Simalungun, "Indikator Kesejahteraan Rakyat Kabupaten Simalungun," vol. 44102004, no. 1209, p. 22, 2019.
- [2] I. G. M. Y. Antara and I. G. P. E. Suryana, "Pengaruh Tingkat Kepadatan Penduduk Terhadap Indeks Pembangunan Manusia di Provinsi Bali," *Media Komun. Geogr.*, vol. 21, no. 1, p. 63, 2020, doi: 10.23887/mkg.v21i1.22958.
- [3] Y. H. Susanti and E. Widodo, "Perbandingan K-Means dan K-Medoids Clustering terhadap Kelayakan Puskesmas di DIY Tahun 2015," *Pros. SI MaNIs (Seminar Nas. Integr. Mat. dan Nilai Islam.*, vol. 1, no. 1, pp. 116–122, 2017.
- [4] R. Rosmini, A. Fadlil, and S. Sunardi, "Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah," *It J. Res. Dev.*, vol. 3, no. 1, pp. 22–31, 2018, doi: 10.25299/itjrd.2018.vol3(1).1773.
- [5] R. M. Sabiq and N. Nurwati, "Pengaruh Kepadatan Penduduk Terhadap Tindakan Kriminal," *J. Kolaborasi Resolusi Konflik*, vol. 3, no. 2, pp. 161–167, 2021.
- [6] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [7] A. N. Khomarudin, "Teknik Data Mining : Algoritma K-Means Clustering," pp. 1–12, 2016.
- [8] A. A. Fajrin and A. Maulana, "Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp- Growth Pada Data Transaksi Penjualan," vol. 05, no. 01, pp. 27–36, 2018.
- [9] T. R. Rivanthio, "Penerapan Metode Clustering K-Means Untuk Pengelompokan Prestasi Mahasiswa Di Politeknik LP3I Bandung," vol. 8, no. 1, pp. 1–13, 2021.
- [10] R. W. Sari and D. Hartama, "Data Mining : Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi," pp. 322–326, 2018.
- [11] D. Triyansyah and D. Fitriana, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *J. Telekomun. dan Komput.*, vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [12] R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018, doi: 10.30865/komik.v2i1.930.
- [13] D. Yanto, L. Probolinggo, C. Loyal, and K. Loyal, "Analisis RFM dan Algoritma K-Means untuk Clustering Loyalitas Customer," vol. 9, no. 1, pp. 0–8, 2019.
- [14] P. V. M., I. Gunawan, B. E. Damanik, I. Parlina, and W. Saputra, "Penerapan Data Mining Menggunakan Algoritma C4.5 Dalam Menentukan Kelayakan Penerimaan Bantuan Bedah Rumah Pada Desa Tiga Dolok," vol. 1, pp. 396–409, 2021.
- [15] W. Muslehatin and M. Ibnu, "Penerapan Naïve Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau," pp. 18–19, 2017.