

# Penerapan Algoritma *K-Means* dalam Pengelompokan Jumlah Penduduk Berdasarkan Kelurahan di Kota Pematangsiantar

Luvita Yolanda Hutabarat <sup>1✉</sup>, Indra Gunawan <sup>2)</sup>, Ika Purnamasari <sup>3)</sup>, M Safii <sup>4)</sup>,  
Widodo Saputra <sup>5)</sup>

<sup>1)2)3)4)5)</sup> Program Studi Teknik Informatika, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

<sup>1)</sup> luvitahutabarat04@gmail.com

<sup>2)</sup> indra@amiktunasbangsa.ac.id

<sup>3)</sup> ikapurnamasari1319@gmail.com

<sup>4)</sup> m.safii@amiktunasbangsa.ac.id

<sup>5)</sup> widodo@amiktunasbangsa.ac.id

**Abstract**— The increase in population in a city or district if not resolved properly and systematically will result in bad impacts. Population growth will affect economic growth and population density. In this study, the population in Pematangsiantar city will be grouped to determine the low, medium and high population so that there is no population density. This study uses data mining techniques using the K-Means Clustering algorithm. This study uses 3 clusters and the results obtained from the calculation of the K-Means algorithm are 29 urban villages with low population clusters, 18 urban villages with moderate population clusters, 6 urban villages with high population clusters.

**Keywords**—Data Mining, K-Means, Clustering, Total Population.

**Intisari**— Pertambahan jumlah penduduk di suatu kota ataupun kabupaten jika tidak diselesaikan dengan baik dan sistematis akan mengakibatkan dampak-dampak yang buruk. Pertambahan jumlah penduduk akan mempengaruhi pertumbuhan ekonomi dan kepadatan penduduk. Pada penelitian ini akan dilakukan pengelompokan jumlah penduduk di kota Pematangsiantar untuk mengetahui jumlah penduduk yang rendah, sedang dan tinggi agar tidak terjadi kepadatan penduduk. Penelitian ini menggunakan teknik data mining dengan menggunakan algoritma *K-Means Clustering*. Penelitian ini memakai 3 *Cluster* dan hasil yang didapatkan dari perhitungan algoritma *K-Means* adalah *Cluster* jumlah penduduk rendah sebanyak 29 kelurahan, *Cluster* jumlah penduduk sedang sebanyak 18 kelurahan, *Cluster* jumlah penduduk tinggi sebanyak 6 kelurahan.

**Kata Kunci**—Data Mining, K-Means, Clustering, Jumlah Penduduk.

## I. PENDAHULUAN

Pada dasarnya penduduk merupakan modal dasar pembangunan, oleh karena itu data statistik kependudukan sepenuhnya diperlukan untuk kepentingan perencanaan pembangunan dengan segala aspeknya. Pertumbuhan

penduduk yang tidak seimbang dengan pertumbuhan kesempatan kerja, mengakibatkan meningkatnya jumlah pengangguran. Pertambahan jumlah penduduk di suatu kota ataupun kabupaten jika tidak diselesaikan dengan baik dan sistematis akan mengakibatkan dampak-dampak yang buruk. Pertambahan jumlah penduduk juga akan mempengaruhi pertumbuhan ekonomi dan kesejahteraan rakyat [1].

Dengan banyaknya pertambahan jumlah penduduk di kota Pematangsiantar, diperlukan adanya pengelompokan jumlah penduduk berdasarkan kelurahan sebagai salah satu upaya untuk mencegah kepadatan penduduk dan dampak-dampak buruk lainnya. Oleh karena itu, dibuatlah pengolahan data yang dapat menyelesaikan permasalahan tersebut yaitu menerapkan teknik data mining menggunakan algoritma *K-Means* terhadap data jumlah penduduk berdasarkan kelurahan di kota Pematangsiantar.

*K-Means* merupakan salah satu metode data *Clustering* non-hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *Cluster*/kelompok. Metode ini mempartisi data ke dalam *Cluster*/kelompok sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu *Cluster* yang sama [2].

Pada penelitian yang dilakukan oleh Sonang dan tim menggunakan algoritma *K-Means* untuk mengelompokkan jumlah penduduk berdasarkan kategori usia. Hasil pengujian yang didapat yaitu kelompok jumlah penduduk tinggi berdasarkan usia ada 3 kecamatan sedangkan kelompok jumlah penduduk rendah berdasarkan usia ada 5 kecamatan [3].

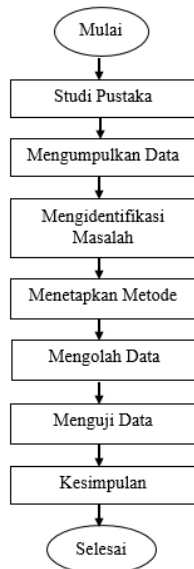
Pada penelitian ini data jumlah penduduk yang digunakan diambil dari Badan Pusat Statistik Pematangsiantar. Data jumlah penduduk yang diperoleh per kelurahan di Kota Pematangsiantar yaitu dari tahun 2015 sampai tahun 2019 yang terdiri dari 53 kelurahan. Data jumlah penduduk yang diperoleh akan dilakukan penelitian dalam mengelompokkan jumlah penduduk berdasarkan kelurahan menggunakan

algoritma *K-Means*. Pengelompokan dibagi menjadi 3 *cluster* yaitu *cluster* jumlah penduduk rendah, *cluster* jumlah penduduk sedang, dan *cluster* jumlah penduduk tinggi. Pengelompokan dilakukan untuk menjadi sebuah solusi agar pemerintah memperhatikan kelompok jumlah penduduk untuk mencegah kepadatan penduduk dan dampak-dampak buruk lainnya supaya kesejahteraan masyarakat lebih terjamin.

II. METODOLOGI

A. Rancangan Penelitian

Rancangan penelitian ini yang dilakukan oleh penulis pertama kali yaitu melakukan observasi dan mengumpulkan data langsung dari Badan Pusat Statistik (BPS) kota Pematangsiantar. Data tersebut akan diolah menggunakan algoritma *K-Means* dan diuji menggunakan *RapidMiner* untuk mengetahui apakah pengujian *RapidMiner* sesuai dengan perhitungan manual menggunakan algoritma *K-Means*. Berikut rancangan penelitian dapat dilihat pada gambar 1.



Gambar 1. Rancangan Penelitian

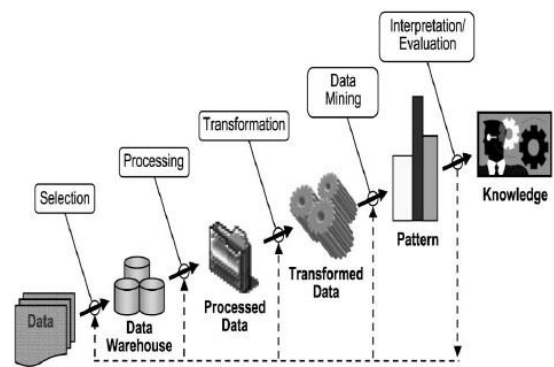
B. Data Mining

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machinelearning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [4]. Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu[5] :

1. Classification Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan.
2. Association Digunakan untuk mengenali kelakuan dari kejadian-kejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian.

3. Clustering Digunakan untuk menganalisis pengelompokan berbeda terhadap data, mirip dengan klasifikasi, namun pengelompokan belum didefinisikan sebelum dijalankannya tool data mining.

“Tahapan yang dilakukan pada proses data mining diawali dari seleksi data dari data sumber ke data target, tahap *pre-processing* untuk memperbaiki kualitas data, transformasi, data mining serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik” [6]. Secara detail dijelaskan sebagai berikut:



Gambar 2. Tahapan Proses Data Mining

C. Clustering

*Clustering* merupakan salah satu teknik dari salah satu fungsionalitas datamining, algoritma *Clustering* merupakan pengelompokan sejumlah data menjadi kelompok-kelompok data tertentu (*Cluster*)[7].

D. RapidMiner

*RapidMiner* merupakan perangkat lunak yang dapat diakses oleh siapa saja dan bersifat terbuka (*open source*). *RapidMiner* ini dijadikan sebuah solusi untuk menganalisa terhadap data *processing*. Pada *RapidMiner* ini digunakan berbagai teknik seperti teknik deskriptif dan prediksi. *RapidMiner* ini menggunakan bahasa Java untuk pengoperasian nya [8].

E. Algoritma *K-Means*

*K-Means* merupakan salah satu algoritma *Clustering* yang masuk dalam kelompok *Unsupervised Learning* yang digunakan untuk membagi data menjadi beberapa kelompok dengan sistem partisi. Algoritma ini menerima masukan berupa data tanpa label kelas. Pada setiap *Cluster* terdapat titik pusat (*Centroid*) yang mempresentasikan *Cluster* tersebut. Secara sederhana algoritma *K-Means* dapat dijelaskan sebagai algoritma data mining yang digunakan untuk menyelesaikan masalah pengelompokan (*Clustering*)[9].

Langkah-langkah melakukan *Clustering* dengan metode *K-Means* adalah sebagai berikut [10]:

1. Tentukan jumlah *Cluster* (k) pada data set
2. Tentukan nilai pusat (*Centroid*) Penentuan nilai *Centroid* pada tahap awal dilakukan secara random,

sedangkan pada tahap iterasi digunakan rumus seperti berikut ini:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Keterangan:

$V_{ij}$  = Centroid rata-rata Cluster ke-i untuk variable ke-j

$N_i$  = Jumlah anggota Cluster ke-i

i,k = Indeks dari Cluster

j = Indeks dari variable

$X_{kj}$  = Nilai data ke-k variable ke-j untuk Cluster tersebut

3. Pada masing-masing record, hitung jarak terdekat dengan Centroid. Jarak Centroid yang digunakan adalah Euclidean Distance, dengan rumus seperti berikut ini:

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Keterangan:

De = Euclidean Distance

i = banyaknya objek

(x, y) = Koordinat objek

(s,t) = Koordinat Centroid

4. Kelompokkan objek berdasarkan jarak ke Centroid terdekat
5. Ulangi langkah ke-2, lakukan iterasi hingga Centroid bernilai optimal

### III. HASIL DAN PEMBAHASAN

#### A. Perhitungan Dengan Algoritma K-Means

Pengelompokan data dilakukan dengan menggunakan K-Means Clustering dengan sampel data jumlah penduduk setiap kelurahan. Untuk mendapatkan hasil dari penelitian dilakukan perhitungan manual untuk proses Clustering menggunakan algoritma K-Means. Proses Clustering dilakukan mulai dari penentuan data yang ingin di Cluster. Berikut adalah data sampel.

TABEL 1. DATA JUMLAH PENDUDUK

No	Kelurahan	Jumlah Penduduk				
		2015	2016	2017	2018	2019
1	Pardamean	3,536	3,484	3,651	3,557	3,589
2	Parhorasan Nauli	3,054	3,011	3,965	3,073	3,102
3	Sukamaju	3,868	3,827	3,859	3,893	3,928
4	Sukamakmur	1,747	1,742	3,521	1,777	1,794
5	Sukaraja	2,663	2,624	2 760	2,678	2,703
6	BP Nauli	2 340	2,935	3,042	2,995	3,022
7	Mekar Nauli	1,888	1,634	1,651	1,669	1,684
8	Dwi Kora	2,329	1,348	2,371	2,393	2,414
9	Proklamasi	1 640	1,654	1,672	1 690	1,707
10	Bantan	11,114	11,209	11,308	11,406	11,496
11	Timbang Galung	3,237	3,264	3,294	3,323	3,35
12	Simarito	6,555	6 610	6 670	6,729	6,785

13	Sipinggol Pinggol	4,578	4,617	4,656	4,689	4,735
14	Banjar	5 220	5,264	5,313	5,362	5,408
15	Teladan	2,452	2,473	2,498	2,522	2,545
16	Kebun Sayur	4,248	4,283	4 310	4,336	4,358
17	Tomuan	9,575	9,656	9,718	9,779	9,832
18	Pahlawan	2,628	2,651	2,692	2,734	2,775
19	Asuhan	4,949	4,991	5,019	5,047	5,072
20	Merdeka	3,508	3,537	3,557	3,577	3,594
21	Pardomuan	4,412	4,449	4,474	4,499	4 520
22	Siopat Suhu	10,882	10,975	11,041	11,107	11,165
23	Toba	2,266	2,286	2,302	2,318	2,333
24	Karo	3,387	3,415	3,436	3,457	3,475
25	Simalungun	2,543	2,565	2,581	2,596	2,609
26	Martimbang	3,616	3,646	3,668	3 690	3,709
27	Kristen	2 220	2,238	2,254	2 270	2,284
28	Aek Nauli	3,827	3 860	3,884	3,908	3,929
29	Melayu	7 650	7,715	7,769	7,822	7 870
30	Martoba	9,148	9,226	9,287	9,348	9,453
31	Baru	6,099	6,150	6,191	6,231	6,267
32	Sukadame	5,394	5 440	5,476	5,511	5,542
33	Kahean	7,312	7,374	7,421	7,467	7,508
34	Sigulang Gulang	6,228	6 280	6,322	6,363	6 400
35	Bane	6,708	6,765	6,811	6,856	6,896
36	Sumber Jaya	5,773	5,821	5,858	5,894	5,926
37	Tambun Nabolon	6 140	6,192	6,265	6 320	6,38
38	Tanjung Tengah	3,277	3,305	3,339	3,373	3,405
39	Pondok Sayur	6,065	6,117	6,164	6 210	6,252
40	Tanjung Pinggir	4,636	4,675	4,711	4,746	4,777
41	Naga Pita	10,026	10,112	10,191	10,268	10,339
42	Naga Pitu	4,549	4,587	4,662	4,657	4,689
43	Gurilla	1,981	1,981	2,021	2,026	2,039
44	Setia Negara	7,241	7,241	7,399	7,496	7,589
45	Bah Kapul	9,772	9,772	9,936	10,016	10,089
46	Bah Sorma	3 360	3 360	3,426	3,464	3 500
47	Bukit Shofa	6,163	6,163	6,256	6,296	6,331
48	Simarimbun	2,246	2,265	2,286	2,307	2,326
49	Tong Marimbun	3,096	2,198	2,219	2 240	3,218
50	Nagahuta	3,045	3,121	3,154	3,187	3,157
51	Nagahuta Timur	2,179	2,116	2,139	2,162	2 260
52	Pematang Marihat	2,098	3,071	3,101	3 130	2,183
53	Marihat Jaya	2,943	2,968	2,998	3,027	3,054

- Menentukan Jumlah k *Cluster*, Pada tahap ini dapat diketahui:  
 Jumlah *Cluster*: 3  
 Jumlah Data: 53
- Menentukan *Centroid* awal secara random, penentuan pusat *Cluster* awal ditentukan secara random yang diambil dari data yang ada dalam *range*.

TABEL 2. CENTROID DATA AWAL

Kelurahan Gurilla	1,981	1,981	2,021	2,026	2,039
Kelurahan Baru	6,099	6,150	6,191	6,231	6,267
Kelurahan Bantan	11,114	11,209	11,308	11,406	11,496

- Menghitung *Centroid* terdekat, berikut adalah contoh perhitungan jumlah penduduk dengan algoritma *K-Means*:

Kel. Pardamean

Data(1,1)=

$$\sqrt{(3,536-1,981)^2 + (3,484-1,981)^2 + (3,651-2,021)^2 + (3,557-2,026)^2 + (3,589-2,039)^2} = 3,475686$$

Data(1,2)=

$$\sqrt{(3,536-6,099)^2 + (3,484-6,150)^2 + (3,651-6,191)^2 + (3,557-6,231)^2 + (3,589-6,267)^2} = 5,869419$$

Data(1,3)=

$$\sqrt{(3,536-11,114)^2 + (3,484-11,209)^2 + (3,651-11,308)^2 + (3,557-11,406)^2 + (3,589-11,496)^2} = 17,31643$$

Kel. Marihata Jaya

Data(53,1)=

$$\sqrt{(2,943-1,981)^2 + (2,968-1,981)^2 + (2,998-2,021)^2 + (3,027-2,026)^2 + (3,054-2,039)^2} = 2,210513$$

Data(53,2)=

$$\sqrt{(2,943-6,099)^2 + (2,968-6,150)^2 + (2,998-6,191)^2 + (3,027-6,231)^2 + (3,054-6,267)^2} = 7,132299$$

Data(53,3)=

$$\sqrt{(2,943-11,114)^2 + (2,968-11,209)^2 + (2,998-11,308)^2 + (3,027-11,406)^2 + (3,054-11,496)^2} = 18,57984$$

Setelah semua data selesai dihitung, maka berikut ini adalah hasil dari perhitungan pada iterasi ke-1:

TABEL 3. HASIL PERHITUNGAN JARAK PUSAT *CLUSTER* ITERASI 1

No	Kelurahan	Jarak Ke <i>Centroid</i>			Jarak Terdekat
		<i>Cluster</i> 1	<i>Cluster</i> 2	<i>Cluster</i> 3	
1	Pardamean	3,475686	5,869419	17,31643	3,475686
2	Parhorasan Nauli	2,866626	6,63857	18,05515	2,866626
3	Sukamaju	4,171419	5,171951	16,61944	4,171419
4	Sukamakmur	1,57604	9,238242	20,61186	1,57604
5	Sukaraja	1,513504	7,831758	19,27913	1,513504
6	BP Nauli	1,996669	7,442058	18,87651	1,996669

7	Mekar Nauli	0,720716	10,02767	21,47421	0,720716
8	Dwi Kora	0,958961	9,024431	20,44511	0,958961
9	Proklamasi	0,753745	10,09615	21,54373	0,753745
10	Bantan	20,79027	11,44773	0	0
11	Timbang Galung	2,871425	6,471322	17,91889	2,871425
12	Simarito	10,42137	1,079493	10,3689	1,079493
13	Sipinggol Pinggol	5,915751	3,427031	14,87454	3,427031
14	Banjar	7,388202	1,954882	13,40209	1,954882
15	Teladan	1,092468	8,250403	19,69797	1,092468
16	Kebun Sayur	5,137302	4,205396	15,65309	4,205396
17	Tomuan	17,22377	7,881108	3,567135	3,567135
18	Pahlawan	1,53646	7,807491	19,25478	1,53646
19	Asuhan	6,721805	2,620925	14,06865	2,620925
20	Merdeka	3,454821	5,887924	17,33562	3,454821
21	Pardomuan	5,503552	3,839179	15,2869	3,839179
22	Siopat Suhu	20,17991	10,83725	0,615492	0,615492
23	Toba	0,651852	8,691066	20,1387	0,651852
24	Karo	3,185165	6,157559	17,60524	3,185165
25	Simalungun	1,27291	8,069923	19,51759	1,27291
26	Martimbang	3,703492	5,639227	17,08691	3,703492
27	Kristen	0,544995	8,797964	20,2456	0,544995
28	Aek Nauli	4,186066	5,156637	16,60432	4,186066
29	Melayu	12,87051	3,527861	7,919898	3,527861
30	Martoba	16,28588	6,943319	4,504564	4,504564
31	Baru	9,342665	0	11,44773	0
32	Sukadame	7,743805	1,598867	13,04658	1,598867
33	Kahean	12,09043	2,747767	8,700061	2,747767
34	Sigulang Gulang	9,635606	0,292942	11,15479	0,292942
35	Bane	10,72822	1,385558	10,06218	1,385558
36	Sumber Jaya	8,597536	0,745155	12,19288	0,745155
37	Tambun Nabolon	9,503897	0,172078	11,28644	0,172078
38	Tanjung Tengah	2,974909	6,367953	17,81545	2,974909
39	Pondok Sayur	9,284677	0,060332	11,50563	0,060332
40	Tanjung Pinggir	6,036373	3,306315	14,75394	3,306315
41	Naga Pita	18,28673	8,944122	2,503704	2,503704
42	Naga Pitu	5,857081	3,48581	14,93331	3,48581
43	Gurilla	0	9,342665	20,79027	0
44	Setia Negara	12,04083	2,702151	8,750896	2,702151
45	Bah Kapul	17,68302	8,340788	3,108065	3,108065
46	Bah Sorma	3,159062	6,184175	17,63147	3,159062
47	Bukit Shofa	9,464018	0,129657	11,32637	0,129657

48	Simarimbun	0,618414	8,724512	20,1721	0,618414
49	Tong Marimbun	1,662936	8,101702	19,50993	1,662936
50	Nagahuta	2,512614	6,831044	18,27849	2,512614
51	Nagahuta Timur	0,372411	8,981708	20,42897	0,372411
52	Pematang Marihat	1,899405	7,831449	19,23783	1,899405
53	Marihat Jaya	2,210513	7,132299	18,57984	2,210513

	Sigulang Gulang, Bane, Sumber Jaya, Tambun Nabolon, Pondok Sayur, Tanjung Pinggir, Naga Pitu, Setia Negara, Bukit Shofa.	
C3	Bantan, Tomuan, Siopat Suhu, Martoba, Naga Pita, Bah Kapul.	6

4. Menghitung *Centroid* baru untuk iterasi berikutnya dengan menghitung rata-rata nilai pada masing-masing *Cluster*. Berikut ini adalah perhitungan *Centroid* baru pada masing-masing *Cluster*:

$$C1_1 = \frac{3,536 + 3,054 + 3,868 + 1,747 + 2,663 + 2,340 + 1,888 + 2,329 + 1,640 + 3,237 + 2,452 + 2,628 + 3,508 + 2,266 + 3,387 + 2,543 + 3,616 + 2,220 + 3,827 + 3,277 + 1,981 + 3,360 + 2,246 + 3,096 + 3,045 + 2,179 + 2,098 + 2,943}{28} = 2,749071$$

$$C1_2 = \frac{3,484 + 3,011 + 3,827 + 1,742 + 2,624 + 2,935 + 1,634 + 1,348 + 1,654 + 3,264 + 2,473 + 2,651 + 3,537 + 2,286 + 3,415 + 2,565 + 3,646 + 2,238 + 3,860 + 3,305 + 1,981 + 3,360 + 2,265 + 2,198 + 3,121 + 2,116 + 3,071 + 2,968}{28} = 2,734964$$

$$C1_3 = \frac{3,651 + 3,965 + 3,859 + 3,521 + 2,760 + 3,042 + 1,651 + 2,371 + 1,672 + 3,294 + 2,498 + 2,692 + 3,557 + 2,302 + 3,436 + 2,581 + 3,668 + 2,254 + 3,884 + 3,339 + 2,021 + 3,426 + 2,286 + 2,219 + 3,154 + 2,139 + 3,101 + 2,998}{28} = 2,905036$$

Jika tahapan iterasi telah mencapai hasil yang sama tanpa dan tidak mengalami perubahan lagi, maka perhitungan dihentikan.

Pada penelitian ini perhitungan dilakukan sampai iterasi ke-3. Setelah perhitungan selesai dilakukan sampai iterasi ke 3 dan telah dikelompokkan, maka hasil akhir posisi *Cluster* dari *Centroid* iterasi ke 3 dapat dilihat pada tabel dibawah ini:

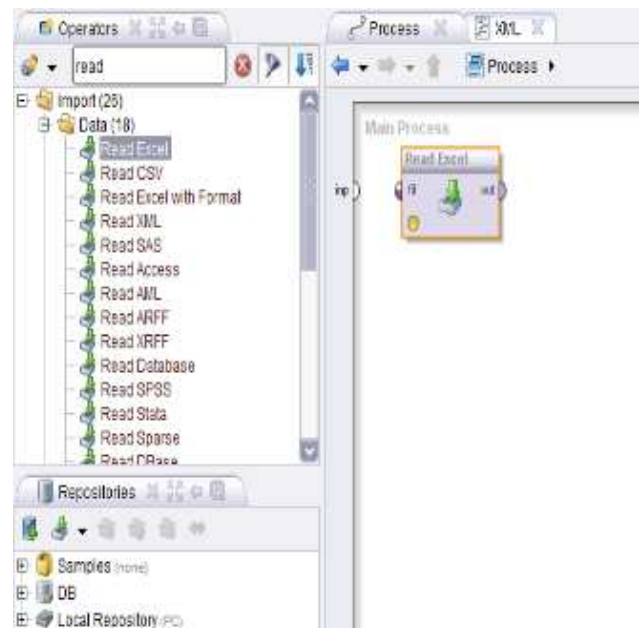
TABEL 4. HASIL PENGELOMPOKAN ITERASI KE 3

Cluster	Kelurahan	Jumlah
C1	Pardamean, Parhorasan Nauli, Sukamaju, Sukamakmur, Sukaraja, BP Nauli, Mekar Nauli, Dwi Kora, Proklamasi, Timbang Galung, Teladan, Kebun Sayur, Pahlawan, Merdeka, Toba, Karo, Simalungun, Martimbang, Kristen, Aek Nauli, Tanjung Tengah, Gurilla, Bah Sorma, Simarimbun, Tong Marimbun, Nagahuta, Nagahuta Timur, Pematang Marihat, Marihat Jaya.	29
C2	Simarito, Sippinggol Pinggol, Banjar, Asuhan, Pardomuan, Melayu, Baru, Sukadame, Kahean,	18

B. Implementasi Pada Aplikasi RapidMiner

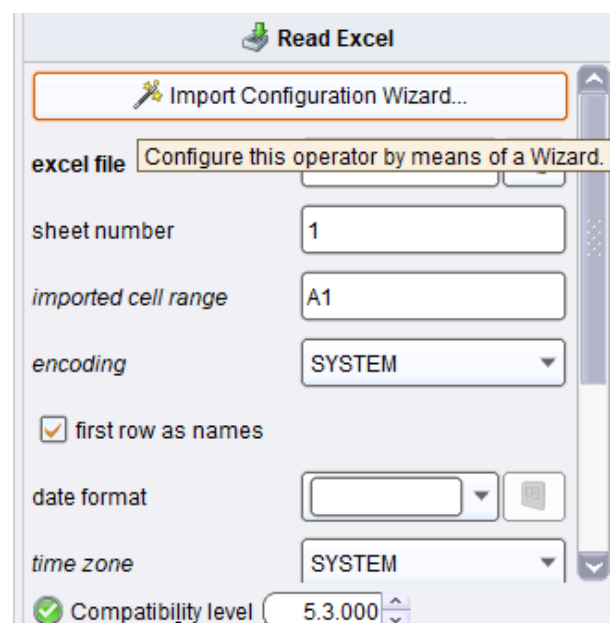
Dalam Pengimplementasian pada aplikasi *RapidMiner* dapat dilihat sebagai berikut:

1. Import Data kedalam *RapidMiner* dalam bentuk Sheet Excel



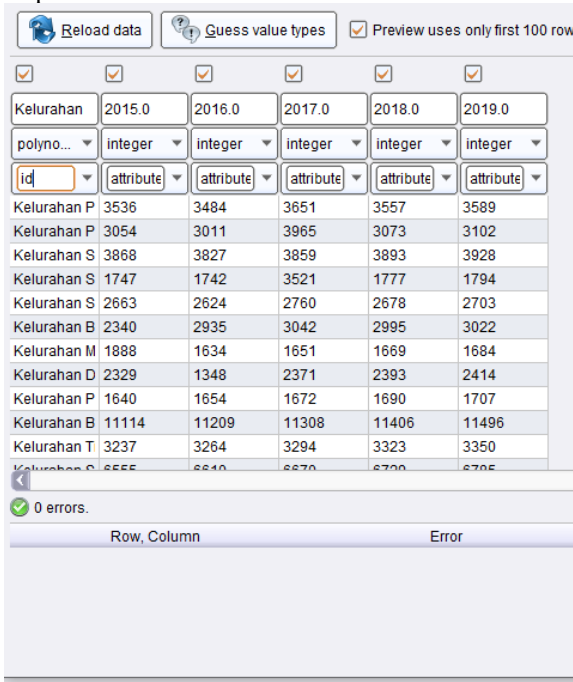
Gambar 3. Import Data

2. Import Configuration Wizard



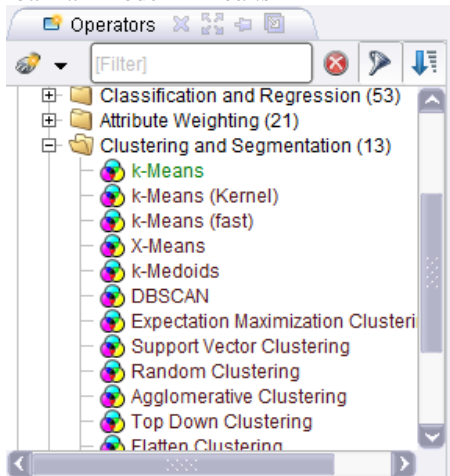
Gambar 4. Konfigurasi Wizard

3. Setelah Configuration Wizard Pilih file, sesuaikan Atribut dan Tipe Data



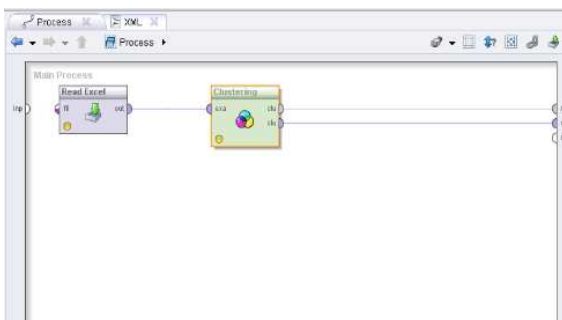
Gambar 5. Penyesuaian Atribut dan Tipe Data

4. Menambahkan Model K-Means



Gambar 6. Import Model K-Means

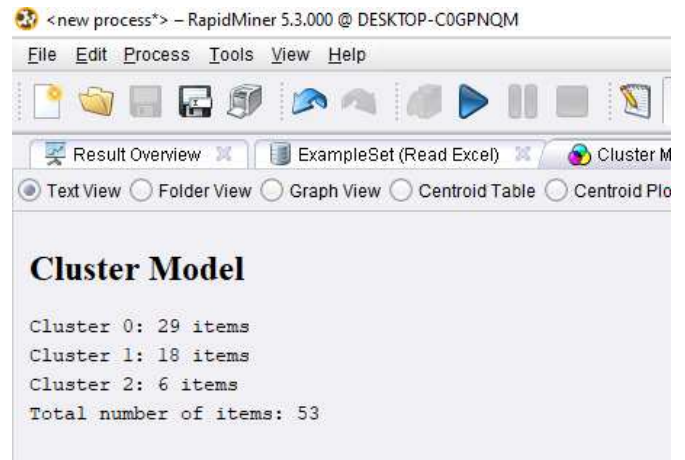
5. Pemrosesan Sistem RapidMiner



Gambar 7. Pemrosesan Data

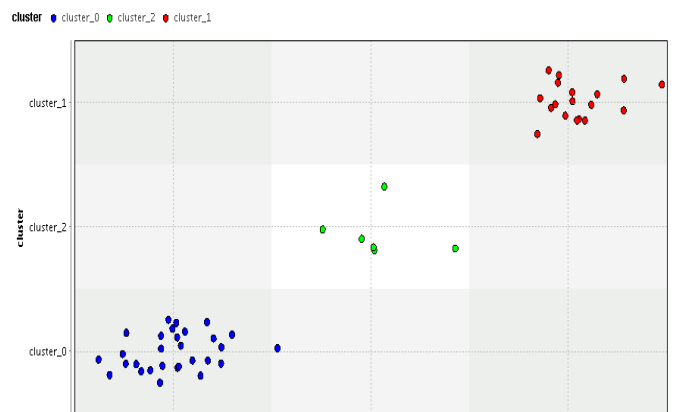
6. Hasil Pengelompokan dalam RapidMiner

Tampilan Text View



Gambar 8. Hasil Tampilan Text View

Tampilan Plot View



Gambar 9. Hasil Tampilan Plot View

Setelah melakukan implementasi menggunakan aplikasi *RapidMiner*, maka hasil yang didapat dari implementasi *RapidMiner* sama dengan hasil dari perhitungan manual *K-Means*.

IV. KESIMPULAN

Berdasarkan pembahasan dan hasil implementasi software *RapidMiner* pada pengelompokan jumlah penduduk berdasarkan kelurahan, maka dapat disimpulkan sebagai berikut: Penerapan *data mining* dengan algoritma *K-Means* dapat diterapkan, dan sumber data yang digunakan pada penelitian ini adalah dari BPS kota Pematangsiantar dengan subjek data jumlah penduduk berdasarkan kelurahan (2015-2019). Jumlah *record* yang digunakan sebanyak 53 kelurahan dengan menghasilkan 3 *Cluster* yaitu *Cluster* rendah sebanyak 29 kelurahan yaitu Pardamean, Parhorasan Nauli, Sukamaju, Sukamakmur, Sukaraja, BP Nauli, Mekar Nauli, Dwi Kora, Proklamasi, Timbang Galung, Teladan, Kebun Sayur, Pahlawan, Merdeka, Toba, Karo, Simalungun, Martimbang, Kristen, Aek Nauli, Tanjung Tongah, Gurilla, Bah Sorma, Simarimbun, Tong Marimbun, Nagahuta, Nagahuta Timur, Pematang Marihat, Marihat Jaya. Dan *Cluster* sedang

sebanyak 18 kelurahan yaitu Simarito, Sipinggol Pinggol, Banjar, Asuhan, Pardomuan, Melayu, Baru, Sukadame, Kahean, Sigulang Gulang, Bane, Sumber Jaya, Tambun Nabolon, Pondok Sayur, Tanjung Pinggir, Naga Pitu, Setia Negara, Bukit Shofa. Sedangkan *Cluster* tinggi sebanyak 6 kelurahan yaitu Bantan, Tomuan, Siopat Suhu, Martoba, Naga Pita, Bah Kapul.

Berdasarkan hasil pengujian *K-Means* untuk kasus jumlah penduduk menggunakan *Tools RapidMiner* versi 5.3 diperoleh hasil yang sama dengan analisis perhitungan algoritma *K-Means*.

#### UCAPAN TERIMA KASIH

Puji dan syukur penulis ucapkan kepada Tuhan yang Maha Esa, karena Berkat dan Kasih-Nya penulis dapat menyelesaikan penelitian ini. Penulis mengucapkan terima kasih kepada STIKOM Tunas Bangsa serta Bapak Indra Gunawan dan Ibu Ika Purnamasari selaku pembimbing atas izin dan dukungan dalam proses penelitian ini. Serta penulis mengucapkan terima kasih kepada Orang tua, keluarga, dan sahabat atas doa, kasih sayang, nasihat, semangat dan dukungan kepada penulis.

#### REFERENSI

- [1] B. P. Statistik and K. Pematangsiantar, *Katalog: 4102002.1273*. 2017.
- [2] S. Ashari, S. Khansa, C. H. M. Surudin, and I. N. Isnainiyah, "KLUSTERING JUMLAH PENDUDUK KOTA BANDUNG BERDASARKAN JENIS KELAMIN PER KECAMATAN PADA TAHUN 2012 DENGAN METODE K-MEANS," *Semin. Nas. Inform. Sist. Inf. Dan Keamanan Siber*, pp. 22–28, 2018.
- [3] S. Sonang, A. T. Purba, and F. O. I. Pardede, "Pengelompokan Jumlah Penduduk Berdasarkan Kategori Usia Dengan Metode K-Means," *J. Tek. Inf. dan Komput.*, vol. 2, no. 2, pp. 166–172, 2019, doi: 10.37600/tekinkom.v2i2.115.
- [4] R. W. Nasution, Suhada, I. O. Kirana, I. Gunawan, and I. P. Sari, "Penerapan Data Mining Untuk Pengelompokan Minat Konsumen Terhadap Pengguna Jasa Pengiriman Pada PT. Jalur Nugraha Ekakurir (JNE) Pematangsiantar," *RESOLUSI Rekayasa Tek. Inform. dan Inf.*, vol. 1, no. 4, pp. 274–281, 2021.
- [5] D. W. Sitohang and A. Rikki, "Implementasi Algoritma K- Means Clustering untuk Mengelompokkan Data Gizi Balita pada Kecamatan Garoga Tapanuli Utara," *KAKIFIKOM (KUMPULAN Artik. KARYA Ilm. Fak. ILMU Komput.* -, vol. 1, no. 2, pp. 80–92, 2019.
- [6] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 3, no. 2, pp. 173–178, 2018, doi: 10.24114/cess.v3i2.9661.
- [7] K. F. Irnanda, A. P. Windarto, I. S. Damanik, and I. Gunawan, "Penerapan K-Means pada Proporsi Individu dengan Keterampilan (Teknologi Informasi dan Komunikasi) TIK Menurut Wilayah," *Semin. Nas. Sains Teknol. Inf.*, no. c, pp. 452–456, 2019.
- [8] Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer," *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 2, pp. 192–198, 2020, doi: 10.24114/cess.v5i2.18519.
- [9] D. Gultom, H. Dame Tampubolon, L. Yolanda Hutabarat, F. R. Ilmi H Zer, and D. Hartama, "Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 146–151, 2020.
- [10] R. A. Margolang, S. R. Andani, and M. R. Lubis, "Implementasi Data Mining dalam Mengelompokkan Rumah Tangga Kumuh di Perkotaan Berdasarkan Provinsi Menggunakan Algoritma K-Means," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, pp. 602–609, 2019, doi: 10.30645/senaris.v1i0.66.