

## Penerapan Data Mining Untuk Pengelompokan Penyebaran Covid-19 Di Sumatera Utara Menggunakan Algoritma K-Means

Yeusi Fani Selfrida Yani Damanik<sup>\*1</sup>, Sumarno<sup>2</sup>, Indra Gunawan<sup>3</sup>, Dedy Hartama<sup>4</sup>, Ika Okta Kirana<sup>5</sup>

<sup>1,2,3,4,5</sup>Teknik Informatika, STIKOM Tunas Bangsa Pematangsiantar, Indonesia  
Email: <sup>1</sup>yeusifaniselfridayanidamanik@gmail.com, <sup>2</sup>sumarno@gmail.com,  
<sup>3</sup>indra@amiktunasbangsa.ac.id, <sup>4</sup>dedyhartama@amiktunasbangsa.ac.id,  
<sup>5</sup>ikaokta@amiktunasbangsa.ac.id

### Abstrak

Virus corona bisa menyebabkan gangguan ringan pada sistem pernapasan, infeksi paru-paru yang berat, hingga kematian. Penyebaran virus Covid-19 menyebar dengan cepat di berbagai daerah khususnya di Sumatera Utara sehingga tidak memungkinkan untuk melaksanakan aktivitas seperti biasanya. Penyebaran Covid-19 di Sumatera Utara perlu dikelompokkan untuk mengetahui daerah mana yang tingkat penyebarannya tinggi, sedang atau rendah. Data yang digunakan pada penelitian ini adalah Data Kasus Covid 19 Di 33 Kabupaten/ Kota Di Sumatera Utara. Metode dalam penelitian ini adalah algoritma K-Means dan selanjutnya data akan diolah menggunakan aplikasi Rapid Miner menjadi 3 cluster yaitu tinggi, sedang dan rendah. Berdasarkan penelitian ini dibentuk sebuah sistem yang digunakan untuk melatih dan menguji data agar mendapatkan hasil pengelompokan penyebaran Covid-19 di Sumatera Utara.

**Kata kunci:** Covid-19, Algoritma K-Means, Aplikasi Rapid Miner, Clustering

### Abstract

Corona virus can cause mild disorders of the respiratory system, severe lung infections, and even death. The spread of the Covid-19 virus spread quickly in various regions, especially in North Sumatra, making it impossible to carry out activities as usual. The spread of Covid-19 in North Sumatra needs to be grouped to find out which areas have a high, medium or low level of spread. The data used in this study is the Covid 19 Case Data in 33 Regencies/Cities in North Sumatra. The method in this study is the K-Means algorithm and then the data will be processed using the Rapid Miner application into 3 clusters, namely high, medium and low. Based on this research, a system was formed that was used to train and test data in order to get the results of grouping the spread of Covid-19 in North Sumatra.

**Keywords:** Covid-19, K-Means Algorithm, Rapid Miner Application, Clustering.

## 1. PENDAHULUAN

Pada awal 2020, dunia dikejutkan dengan mewabahnya pneumonia baru yang bermula dari Wuhan, Provinsi Hubei yang kemudian menyebar dengan cepat ke lebih dari 190 negara dan teritori. Wabah ini bernama Coronavirus disease 2019 (COVID-19) yang disebabkan oleh Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). Penyebaran penyakit ini telah memberikan dampak luas secara sosial dan ekonomi.[1] Salah satu dampak pandemi Coronavirus 2019–20 ialah terhadap pendidikan di seluruh dunia, yang mengarah kepada penutupan luas sekolah, madrasah, universitas, dan pondok pesantren.[2]

Penyebaran virus Covid-19 menyebar dengan cepat sehingga tidak memungkinkan untuk melaksanakan aktivitas seperti biasanya. Penyebaran Covid-19 penting dibahas agar dapat diketahui daerah-daerah yang terjangkit virus Covid-19 khususnya di Sumatera Utara sehingga dapat memutus mata rantai penyebaran Covid-19. Penyebaran Covid-19 di Sumatera Utara perlu dikelompokkan untuk mengetahui daerah mana yang tingkat penyebarannya tinggi, sedang atau rendah.

Pada penelitian ini algoritma yang digunakan dalam pengelompokan penyebaran virus Covid-19 adalah algoritma K-Means. Algoritma ini dipilih karena sudah teruji dan sering digunakan peneliti untuk melakukan pengelompokan. K-Means merupakan algoritma data clustering non hirarki yang berusaha mempartisi data yang ada dalam bentuk satu atau lebih cluster/kelompok. Algoritma ini mempartisi data dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan dalam kelompok yang lain. [3]

Penelitian yang dilakukan membuat penelitian tentang Coronavirus Disease 2019 : Tinjauan Literatur Terkini. Penelitian ini membahas tentang Coronavirus atau Covid-19 dan penyebarannya. Hasil dari penelitian ini menyatakan bahwa Coronavirus harus diwaspadai karena penularan yang relatif cepat, memiliki tingkat mortalitas yang tidak dapat diabaikan, dan belum adanya terapi definitif.

Penelitian yang dilakukan membahas tentang perilaku masyarakat dalam menghadapi pandemi virus Corona (Covid-19) dan kiat menjaga kesejahteraan jiwa. Hasil dari penelitian ini menunjukkan bahwa perilaku yang ditampilkan oleh orang yang tidak mematuhi himbauan pemerintah didasari oleh bias kognitif.

## 2. TINJAUAN PUSTAKA

### 2.1. Defenisi Dan Konsep Data Mining

Data Mining merupakan teknik pengolahan data dalam jumlah besar untuk pengelompokan[4] Beberapa defenisi Data Mining dari berbagai kutipan yaitu :

- a. Menurut Larose dalam jurnal[3] Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar.
- b. Menurut Afrisawati dalam jurnal [5] Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dalam database. Data Mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan machine learning untuuk mengekstrasi dan mengidentifikasi infromasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.
- c. Menurut Turban dalam jurnal [6] pengertian data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar

Jadi dapat disimpulkan bahwa, data mining adalah sebuah proses untuk menemukan pengetahuan yang belum diketahui yang menggunakan teknik statistik, matematika, kecerdasan buatan serta machine learning yang tersimpan dalam database.

Hal penting yang terkait di dalam data mining adalah:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Data mining merupakan proses penambangan informasi dari besarnya data yang ada dalam basis data. Algoritma Clustering K-Means dirasa cocok sebagai proses pengelompokan data yang sangat besar.[7]

Istilah data mining merupakan bagian dari proses Knowledge Discovery in Database (KDD) yang mentransformasi data mentah menjadi lebih berguna dan bermanfaat.

1. Data Cleansing  
Data Cleansing dikenal sebagai sebuah fase dimana data tidak lengkap, mangandung data yang error dan tidak konsisten akan dibuang dari koleksi data, sehingga data yang telah relevan dapat digunakan untuk penggalian pengetahuan (discovery knowledge).
2. Data Integration

Pada tahap ini terjadi integrasi data, dimana sumber data yang berulang (multiple data) dan file yang berulang (multiple file) dapat dikombinasikan dan digabungkan kedalam suatu sumber.

3. Selection

Pada tahap seleksi ini, data yang relevan terhadap analisis dapat dipilih dan diterima dari koleksi data yang ada.

4. Data Transformation

Tahap ini dikenal sebagai data consolidation yang artinya data yang terpilih akan di transformasikan kedalam bentuk yang cocok untuk prosedur penggalian (meaning procedure) dengan cara melakukan normalisasi dan agregasi data.

5. Data Mining

Tahap ini adalah tahap yang paling penting karena menggunakan teknik yang diaplikasikan untuk mengekstrak pola-pola potensial yang berguna dan bermanfaat.

6. Pattern Evaluation

Pada tahap ini pola-pola yang menarik dengan jelas mempresentasikan pengetahuan yang telah diidentifikasi berdasarkan measure yang telah diberikan.

7. Knowledge Representation

Tahap ini merupakan tahap terakhir dimana pengetahuan yang telah ditemukan secara visual ditampilkan kepada user. Tahap ini menggunakan teknik visualisasi untuk membantu user dalam mengerti dan menginterpretasikan hasil dari data mining.

Tahap ini merupakan tahap terakhir dimana pengetahuan yang telah ditemukan secara visual ditampilkan kepada user. Tahap ini menggunakan teknik visualisasi untuk membantu user dalam mengerti dan menginterpretasikan hasil dari data mining.

Data mining berfungsi untuk mengidentifikasi fakta yang disarankan berdasarkan penyaringan melalui data untuk menjelajahi pola atau anomali data. Fungsi data mining adalah sebagai berikut:

1. Asosiasi (Association)

Digunakan untuk mengenali kelakuan dari kejadian - kejadian khusus atau proses dimana hubungan asosiasi muncul pada setiap kejadian. Salah satu contohnya Market Basket Analysis, yaitu salah satu algoritma asosiasi yang menganalisa kemungkinan pelanggan untuk membeli beberapa item secara bersamaan.

2. Pengklusteran (clustering)

Pengklusteran merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Contoh pengklusteran dalam bisnis dan penelitian adalah mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari satu suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.

3. Prediksi (prediction)

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Contoh prediksi bisnis dan penelitian adalah Prediksi harga beras dalam tiga bulan yang akan datang.

4. Estimasi (estimation)

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik dari pada kearah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Contoh lain yaitu estimasi nilai indeks prestasi kumulatif

mahasiswa program pasca sarjana dengan melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana.

#### 5. Klasifikasi (classification)

Klasifikasi adalah fungsi pembelajaran yang memetakan (mengklasifikasi) sebuah unsur (item) data ke dalam salah satu dari beberapa kelas yang sudah didefinisikan. Contoh lain klasifikasi dalam bisnis dan penelitian adalah mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa.

### 2.2. K-Means

Secara umum pengertian algoritma K-Means dijelaskan secara berbeda antara yang satu dengan yang lainnya namun tetap memiliki arti dan makna yang sama.

Menurut Irwansyah dalam jurnal [8] K-Means merupakan sebuah algoritma clustering pada data mining untuk dapat menghasilkan kelompok dari data yang jumlahnya banyak dengan algoritma partisi yang berbasis titik dengan waktu komputasi yang cepat dan efisien.

K-Means merupakan algoritma clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numeric.[8]

Sedangkan menurut [9], K-Means merupakan algoritma pengklasteran secara partitioning yang memisahkan data ke dalam kelompok yang berbeda. Dengan partitioning secara interaktif, K-Means mampu meminimalkan rata-rata jarak setiap data ke klasternya.

Dari beberapa defenisi tentang algoritma K-Means di atas dapat dikemukakan bahwa algoritma K-Means adalah suatu algoritma clustering yang dapat mengelompokkan data dengan cara partisi dan mampu meminimalkan rata-rata jarak setiap data ke klasternya.

### 2.3. Clustering

Clustering adalah sebuah proses partisi satu set objek data menjadi suatu himpunan bagian yang dikenal dengan sebutan cluster. Objek yang di dalam cluster memiliki kemiripan karakteristik antar satu dengan yang lainnya. Partisi dilakukan dengan suatu algoritma clustering. Clustering sangat berguna dan bisa menemukan group atau kelompok yang tidak dikenal dalam data. Clustering banyak digunakan dalam berbagai aplikasi seperti pada business intelligence, pengenalan pola citra, web search, bidang ilmu biologi, dan untuk keamanan (security). Clustering memiliki banyak kegunaan dan manfaat, yaitu:

1. Clustering adalah algoritma segmentasi data yang sangat bermanfaat dalam memprediksi dan menganalisa masalah bisnis. Contohnya adalah segmentasi pasar, marketing dan pemetaan zonasi wilayah.
2. Clustering dapat mengidentifikasi obyek dalam berbagai bidang seperti computer vision dan image processing.

Dalam clustering terdapat empat tipe data, yaitu:

1. Variabel berskala interval
2. Variabel biner
3. Variabel nominal, ordinal, dan rasio
4. Variabel dengan tipe lainnya

Algoritma clustering juga harus dapat mengukur kemampuannya dalam usaha untuk menemukan suatu pola tersembunyi pada data yang sedang diteliti. Ada beberapa algoritma yang dapat digunakan dalam mengukur nilai kesamaan antar objek yang dibandingkan. Salah satunya adalah dengan Weighted Euclidean Distance yaitu menghitung jarak dua buah point dengan mengetahui nilai dari masing-masing atribut pada kedua point tersebut. Secara umum algoritma clustering dibagi menjadi dua yaitu

#### 1. Hierarchical clustering

Pada hierarchical clustering, data akan dikelompokkan melalui suatu bagan yang berupa hirarki, dimana terdapat penggabungan dua grup yang terdekat disetiap iterasinya ataupun pembagian dari seluruh set data kedalam cluster. Contoh algoritma hierarchical clustering yaitu : Single Linkage, Complete Linkage, Average Linkage, Average Group Linkage.

## 2. Partitional Clustering

Partitional clustering adalah data dikelompokkan kedalam sejumlah cluster tanpa adanya struktur hirarki antara satu dengan yang lainnya. Pada algoritma ini, setiap cluster memiliki titik pusat cluster (centroid) dan secara umum algoritma ini memiliki tujuan untuk meminimumkan jarak (dissimilarity) dari seluruh data ke pusat cluster masing-masing. Contoh algoritma ini yaitu: K-Means, Fuzzy K-Means dan Mixture Modelling.

## 2.4. Covid-19

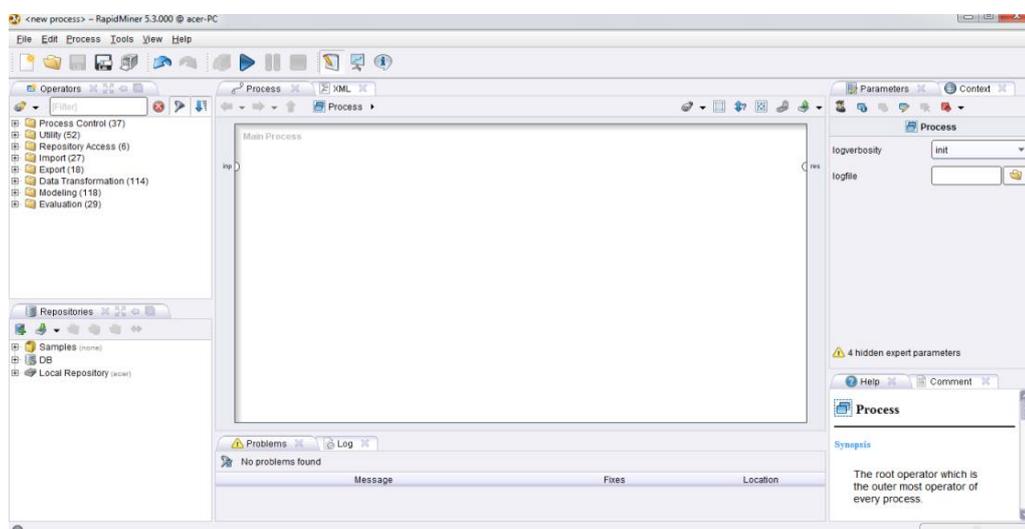
Coronaviruses (CoV) merupakan bagian dari keluarga virus yang menyebabkan penyakit mulai dari flu hingga penyakit yang lebih berat seperti Middle East Respiratory Syndrome (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS-CoV). Penyakit yang disebabkan virus corona, atau dikenal dengan COVID-19, adalah jenis baru yang ditemukan pada tahun 2019 dan belum pernah diidentifikasi menyerang manusia sebelumnya (World Health Organization, 2019).[10]

Adapun cara yang dapat dilakukan adalah sebagai berikut:

1. Rajin mencuci tangan dengan sabun dan air selama 20 detik hingga bersih
2. Jangan menyentuh wajah, hidung, atau mulut saat tangan dalam keadaan kotor atau belum dicuci.
3. Hindari kontak langsung dengan orang yang sakit
4. Hindari menyentuh hewan atau unggah liar
5. Membersihkan atau mensterilkan permukaan benda yang sering digunakan
6. Tutup hidung dan mulut saat bersin atau batuk menggunakan tisu, lalu buanglah tisu dan segera mencuci tangan hingga bersih
7. Tetap berada di rumah dan menghindari keramaian
8. Gunakan masker dan segera berobat ke fasilitas kesehatan terdekat jika mengalami gejala penyakit saluran pernafasan.
9. Perkuat sistem kekebalan tubuh dengan mengkonsumsi makanan sehat dan bila perlu konsumsi vitamin dan suplemen

## 2.5. Rapid Miner

Secara sederhana Rapid Miner merupakan suatu aplikasi yang diciptakan untuk mengolah data menjadi informasi yang lebih bermanfaat dan berguna. Rapid Miner merupakan perangkat lunak yang bersifat terbuka (open source). Rapid Miner adalah software yang berdiri sendiri untuk menganalisa data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. Rapid Miner ditulis menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi. Berikut adalah gambar tampilan rapid miner bisa dilihat pada gambar 1.



Gambar 1. Tampilan *Rapid Miner* versi 5.3

Gambar 1 menjelaskan tentang tampilan Rapid Miner Studio versi 5.3 yang akan digunakan peneliti.

Beberapa fitur yang terdapat pada Rapid Miner yaitu:

- a. Banyaknya algoritma data mining, seperti decision tree dan self-organization map.
- b. Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, tree chart dan 3D Scatter plots.
- c. Banyaknya variasi plugin, seperti text plugin untuk melakukan analisis teks.
- d. Menyediakan prosedur data mining, machine learning termasuk : ETL (extraction, transformation, loading), data processing, visualisasi, modeling dan evaluasi.
- e. Proses data mining tersusun atas operator-operator yang nestable, dideskripsikan dengan XML, dibuat dengan GUI.
- f. Mengintegrasikan proyek data mining Weka dan statistika.

## 2.6. Penelitian Terkait Dan Kontribusi Riset

Berikut ini adalah beberapa penelitian sebelumnya mengenai penggunaan Algoritma K-Means maupun yang berhubungan dengan Covid-19 dirangkum oleh penulis yaitu:

[11] menyampaikan hasil risetnya tentang penerapan RapidMiner dengan K-Means cluster pada daerah terjangkit Demam Berdarah Dengue (DBD) berdasarkan provinsi. Proses iterasi pada penelitian ini berlangsung sebanyak 6 kali sehingga diperoleh hasil penilaian dalam mengelompokkan daerah yang terjangkit DBD berdasarkan provinsi.

[12] dalam risetnya menjelaskan tentang penggunaan algoritma K-Means dalam penelitiannya tentang tehnik data mining dalam mengelompokkan kasus pneumonia pada balita berdasarkan provinsi di Indonesia. Jumlah record yang digunakan dalam penelitian ini sebanyak 34 provinsi yang dibagi dalam 2 cluster yakni cluster tinggi dan cluster rendah. Proses implementasi pada penelitian ini menggunakan RapidMiner 5.3 untuk membantu menemukan nilai yang akurat.

[13] dalam penelitiannya menjelaskan bahwa penelitian ini menggunakan algoritma analisis isi dengan tujuan untuk memberikan edukasi kepada masyarakat dalam menghadapi bencana global dan pencegahannya. Hasil penelitian menunjukkan bahwa pemerintah banyak melakukan strategi komunikasi kepada masyarakat wilayahnya masing-masing untuk mencegah penyebaran Covid-19.

Penelitian ini memberikan kontribusi yakni daerah yang ada di Sumatera Utara dapat dikelompokkan kedalam daerah yang tinggi dan rendah tingkat penyebarannya dengan memanfaatkan algoritma K-Means yang bertujuan untuk mengetahui penyebaran Covid-19 dan membantu masyarakat dalam menghindari daerah yang masuk dalam daerah tinggi tingkat penyebarannya agar penyebaran Covid-19 semakin berkurang di Sumatera Utara.

## 3. METODE PENELITIAN

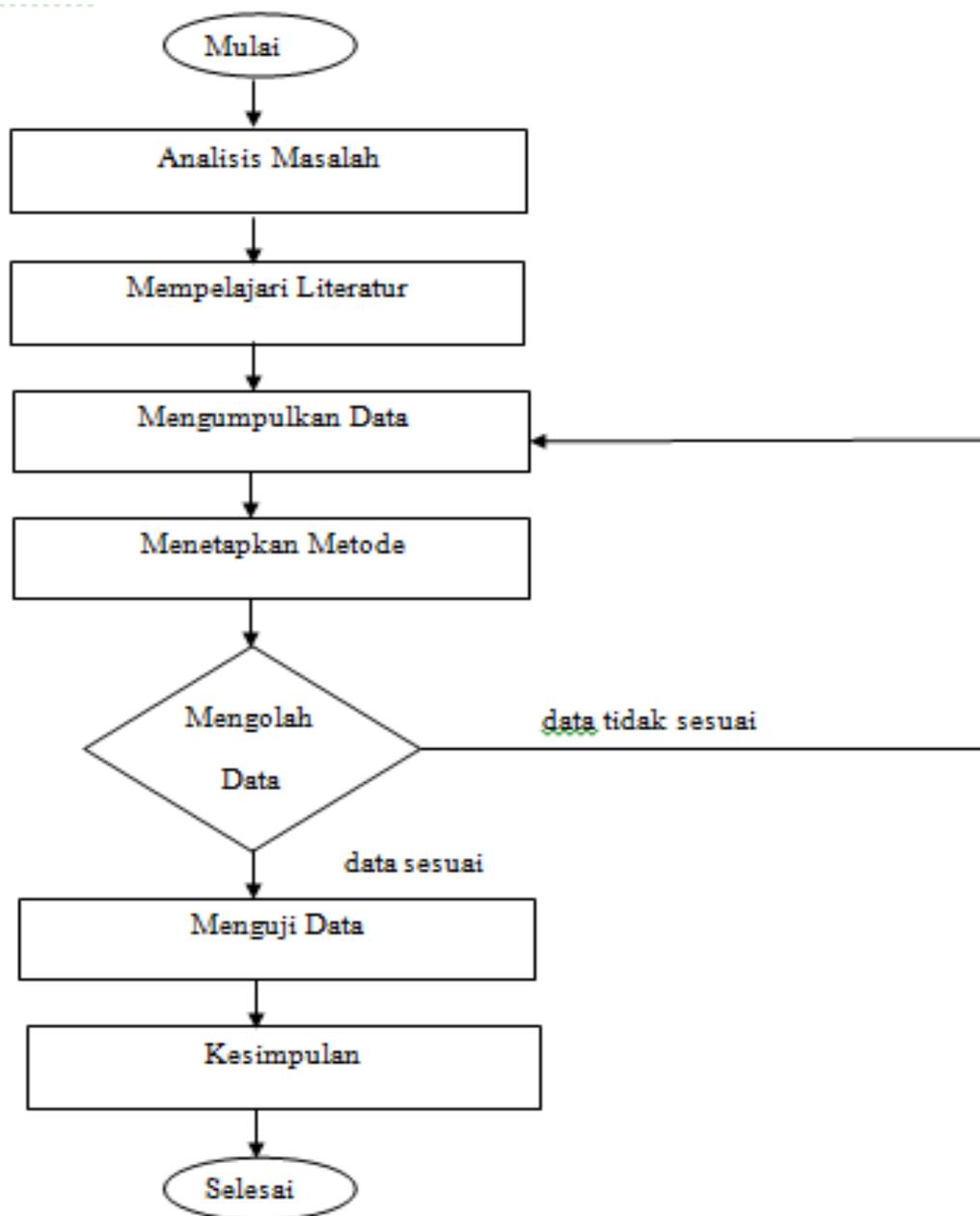
Penelitian ini menguraikan tujuan, model cara ilmiah untuk mengumpulkan data, dan menganalisis data untuk kegunaan tertentu.

### 3.1. Rancangan Penelitian

Jenis data yang dikumpulkan dalam penelitian ini berupa data sekunder. Dalam penelitian ini yang menjadi sumber data sekunder adalah artikel, jurnal serta situs internet yang berhubungan dengan penelitian. Data yang dikumpulkan yaitu Data Kasus Covid-19 di Sumatera Utara.

Adapun atribut yang akan dijadikan prediktor untuk menghasilkan atribut target, yaitu (1) Pasien Dalam Pengawasan (PDP) (2) Positif (3) Sembuh (4) Meninggal. Pada tahap selanjutnya yaitu data ditransformasi kedalam format Microsoft Excel, yang kemudian akan di proses menggunakan aplikasi perangkat lunak Rapid Miner.

Algoritma penelitian pada skripsi ini dapat dideskripsikan sesuai gambar 2 berikut ini :



Gambar 2. Rancangan Penelitian

### 3.2. Prosedure Pengumpulan Data

Algoritma yang penelitian yang digunakan pada pengumpulan data penyebaran Covid-19 di Sumatera Utara adalah menggunakan algoritma studi pustaka. Penelitian ini mengumpulkan data yang relevan dari artikel ilmiah, berita maupun kredibel lainnya yang terkait dengan topik penelitian.

### 3.3. Analisis Data

Data yang digunakan pada penelitian ini menggunakan sumber data yang terekam di situs Pemerintah Provinsi Sumatera Utara dengan alamat url <http://covid19.sumutprov.go.id/>. Data yang digunakan adalah data bulan Juni 2020.

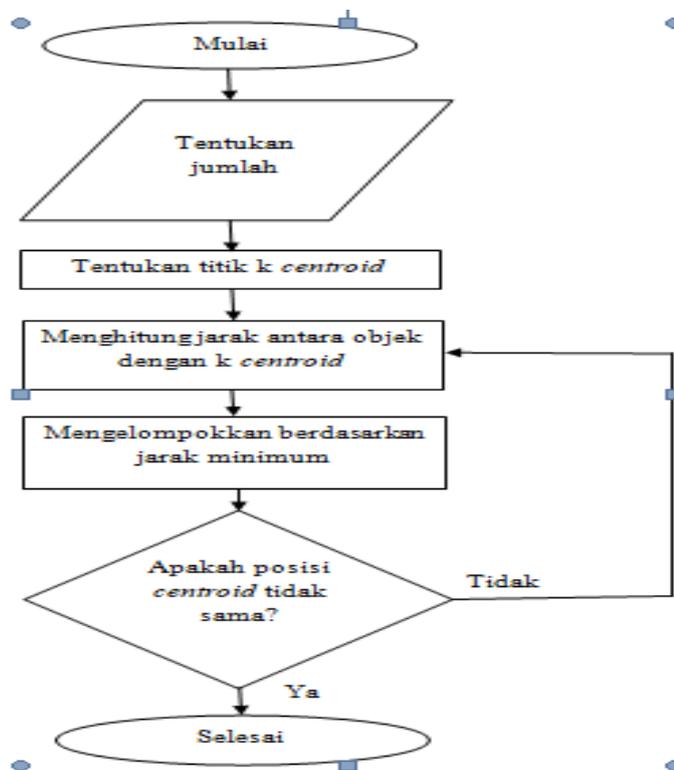
Berikut adalah data yang diolah menggunakan Microsoft Excel dapat dilihat pada Tabel 1

Tabel 1. Data Kasus Covid-19 Di 33 Kabupaten/Kota Di Sumatera Utara

N O	KABUPATEN/ KOTA	JUNI				JULI				AGUSTUS			
		PD P	POSI TIF	SEMB UH	MENING GAL	PD P	POSI TIF	SEMB UH	MENING GAL	PD P	POSI TIF	SEMB UH	MENING GAL
1	KAB. ASAHAN	54	262	64	37	76	894	244	149	8	1489	818	158
2	KAB. BATUBARA	0	54	0	0	47	861	287	0	11	1430	1087	0
3	KAB. DAIRI	23	48	30	0	10	143	41	0	4	451	164	42
4	KAB. DELI	68				12				14			
4	DERDANG	5	3523	981	337	15	11128	3395	788	85	21058	10786	1341
5	KAB. HUMBANG												
5	HASUNDUTAN	12	15	0	0	19	80	39	0	32	221	163	0
6	KAB. KARO	43	327	69	0	67	1117	397	76	78	2572	965	235
7	KAB. LABUHANBATU	11	32	30	0	20	127	55	2	7	391	183	31
8	KAB. LABUHANBATU SELATAN	2	8	0	0	21	79	0	0	0	160	86	0
9	KAB. LABUHANBATU UTARA	28	60	0	60	0	62	0	62	11	203	19	66
10	KAB. LANGKAT	13	223	48	63	18	831	167	92	3	1596	717	142
11	KAB. MANDAILING	25	45	22	0	7	60	60	0	73	104	68	0
12	KAB. NIAS	0	0	0	0	38	0	0	0	8	18	0	0
13	KAB. NIAS BARAT	0	0	0	0	0	0	0	0	0	2	0	0
14	KAB. NIAS SELATAN	0	0	0	0	0	23	0	0	18	37	18	2
15	KAB. NIAS UTARA	2	0	0	0	5	3	0	0	47	48	16	0
16	KAB. PADANG LAWAS	5	34	0	8	0	84	23	31	0	126	33	37
17	KAB. PADANG LAWAS UTARA	57	22	3	0	15	46	31	0	4	92	55	0
18	KAB. PAKPAK BHARAT	0	0	0	0	0	0	0	0	0	2	0	0
19	KAB. SAMOSIR	0	8	0	0	25	60	35	2	26	88	71	0
20	KAB. SERDANG BEDAGAI	62	394	142	30	81	906	414	31	16	2745	1150	71
21	KAB. SIMALUNGUN	50				72				13			
21	SIMALUNGUN	9	1325	369	31	8	3140	1309	128	14	4388	2997	237
22	KAB. TAPANULI SELATAN	0	31	13	0	14	93	41	0	19	187	157	2
23	KAB. TAPANULI TENGAH	5	13	0	0	26	164	14	39	22	1105	751	62
24	KAB. TAPANULI UTARA	47	158	92	0	31	260	164	0	35	441	301	0
25	KAB. TOBA	31	60	60	0	38	290	125	19	47	578	365	31
26	KOTA PEMATANGSIANTAR	26				77				18			
26	PEMATANGSIANTAR	5	1254	354	90	6	3806	1212	132	02	6186	3513	191
27	KOTA BINJAI	43	407	79	27	72	1291	508	96	11	2861	1459	266
28	KOTA GUNUNGSITOLI	0	12	0	0	0	34	21	0	6	240	31	18
29	KOTA MEDAN	30				62				92			
29	KOTA PADANG SIDEMPUAN	25	18539	4804	1164	04	51676	12729	2483	53	96256	46244	4166
30	KOTA SIBOLGA	0	74	0	30	19	168	71	41	0	297	178	80
31	KOTA TANJUNGPALAI	5	5	0	0	12	263	0	8	31	513	208	49
32	KOTA TEBINGTINGGI	13	77	0	49	0	190	2	62	0	310	109	62
33	KOTA TEBINGTINGGI	27	110	30	0	38	528	108	28	51	1550	654	99

### 3.4. Instrumen Penelitian

Dalam penelitian ini penulis melakukan beberapa proses pengumpulan data dan pengolahan data dengan cara memasukkan data kemudian memproses dengan algoritma K-Means clustering. Sebelum diproses data terlebih dahulu di seleksi menggunakan aplikasi Rapid Miner. Flowchart K-Means dapat dilihat pada Gambar 3 dibawah ini.



Gambar 3. Flowchart K-Means

### 3.5. Pemodelan Algoritma

K-Means clustering merupakan algoritma yang dapat menentukan cluster data dengan efektif yang mempartisi N objek ke dalam K kelompok (cluster) berdasarkan nilai rata-rata (means) terdekat.

Ada dua jenis clustering yang sering digunakan para peneliti untuk mengelompokan data yaitu : Hierarchical dan Non-Hierarchical. K-Means masuk dalam kelompok algoritma data clustering Non-Hierarchical atau Partitional Clustering. Beberapa tahapan dalam K-Means clustering adalah sebagai berikut:

1. Tentukan beberapa banyak jumlah k (cluster).
2. Tentukan beberapa record yang akan dijadikan pusat cluster.
3. Temukan pusat cluster terdekat pada setiap recordnya. Berikut adalah persamaan yang sering digunakan dalam Euclidean :  
Dimana  $x=x_1, x_2, x_3, \dots, x_m$  dan  $y=y_1, y_2, y_3, \dots, y_m$ , dan m menyatakan banyaknya nilai atribut dari 2 buah record.
4. Tentukan cluster terdekat untuk setiap data dengan membandingkan nilai jarak terdekat, lalu perbaharui nilai pusat clusternya.
5. Ulangi langkah 3 sampai 5 hingga tidak ada record yang berpindah cluster atau convergen.

## 4. HASIL DAN PEMBAHASAN

Pada penelitian ini data yang digunakan adalah jumlah penyebaran Covid-19 di Sumatera Utara pada bulan Juni – Agustus 2020 yang diperoleh dari website Media Center Gugus Tugas Covid-19

SUMUT. Dalam melakukan clustering, data yang diperoleh akan dihitung terlebih dahulu. Penelitian ini menggunakan aplikasi RapidMiner 5.3 sebagai pengujian data yang di analisa.

Berikut adalah rangkaian perhitungan manual proses Algoritma K-Means Clustering.

### 1. Menentukan Data Yang Akan Di Cluster

Sampel data yang akan digunakan dalam proses clustering adalah data penyebaran Covid-19 di Sumatera Utara. Berikut daftar tabel penyebaran Covid-19 di Sumatera Utara dapat dilihat pada tabel 2.

Tabel 2. penyebaran Covid-19 di Sumatera Utara

NO	KABUPATEN/KOTA	JUNI				JULI				AGUSTUS			
		PDP	POSITIF	SEMBUH	MENINGGAL	PDP	POSITIF	SEMBUH	MENINGGAL	PDP	POSITIF	SEMBUH	MENINGGAL
1	KAB. ASAHAN	54	262	64	37	76	894	244	149	8	1489	818	158
2	KAB. BATUBARA	0	54	0	0	47	861	287	0	114	1430	1087	0
3	KAB. DAIRI	23	48	30	0	10	143	41	0	97	451	164	42
4	KAB. DELI DERDANG	685	3523	981	337	1215	11128	3395	788	1485	21058	10786	1341
5	KAB. HUMBANG	12	15	0	0	19	80	39	0	32	221	163	0
6	KAB. HASUNDUTAN	43	327	69	0	67	1117	397	76	787	2572	965	235
7	KAB. KARO	11	32	30	0	20	127	55	2	50	391	183	31
8	KAB. LABUHANBATU	2	8	0	0	21	79	0	0	0	160	86	0
9	KAB. LABUHANBATU SELATAN	28	60	0	60	0	62	0	62	113	203	19	66
10	KAB. LANGKAT	13	223	48	63	187	831	167	92	730	1596	717	142
11	KAB. MANDAILING	25	45	22	0	38	60	60	0	8	104	68	0
12	KAB. NIAS	0	0	0	0	0	0	0	0	0	18	0	0
13	KAB. NIAS BARAT	0	0	0	0	0	0	0	0	18	2	0	0
14	KAB. NIAS SELATAN	0	0	0	0	5	23	0	0	47	37	18	2
15	KAB. NIAS UTARA	2	0	0	0	0	3	0	0	0	48	16	0
16	KAB. PADANG LAWAS	5	34	0	8	0	84	23	31	0	126	33	37
17	KAB. PADANG LAWAS UTARA	57	22	3	0	15	46	31	0	4	92	55	0
18	KAB. PAKPAK BHARAT	0	0	0	0	0	0	0	0	0	2	0	0
19	KAB. SAMOSIR	0	8	0	0	25	60	35	2	26	88	71	0
20	KAB. SERDANG BEDAGAI	62	394	142	30	81	906	414	31	160	2745	1150	71
21	KAB. SIMALUNGUN	509	1325	369	31	728	3140	1309	128	1314	4388	2997	237
22	KAB. TAPANULI SELATAN	0	31	13	0	14	93	41	0	19	187	157	2
23	KAB. TAPANULI TENGAH	5	13	0	0	26	164	14	39	22	1105	751	62
24	KAB. TAPANULI UTARA	47	158	92	0	31	260	164	0	35	441	301	0
25	KAB. TOBA	31	60	60	0	38	290	125	19	47	578	365	31
26	KOTA PEMATANGSIANTAR	265	1254	354	90	776	3806	1212	132	1802	6186	3513	191
27	KOTA BINJAI	43	407	79	27	72	1291	508	96	111	2861	1459	266
28	KOTA GUNUNGSITOLI	0	12	0	0	0	34	21	0	6	240	31	18
29	KOTA MEDAN	3025	18539	4804	1164	6204	51676	12729	2483	9253	96256	46244	4166
30	KOTA PADANG SIDEMPUAN	0	74	0	30	19	168	71	41	0	297	178	80
31	KOTA SIBOLGA	5	5	0	0	12	263	0	8	316	513	208	49
32	KOTA TANJUNGPALAI	13	77	0	49	0	190	2	62	0	310	109	62
33	KOTA TEBINGTINGGI	27	110	30	0	38	528	108	28	513	1550	654	99

### 2. Menentukan Nilai k Jumlah Cluster

Jumlah cluster yang akan dibentuk adalah 3 cluster yaitu cluster tinggi (C1) , cluster sedang (C2) dan cluster rendah (C3).

3. Menentukan Nilai Centroid Pusat (Pusat Cluster)

Penentuan pusat cluster awal ditetapkan secara acak atau random yang diambil dari data yang ada dalam range. Adapun nilai untuk cluster rendah (C1) diambil dari nilai rendah yang terdapat pada Tabel 4.1 dan nilai untuk cluster sedang (C2) diambil dari nilai sedang dan cluster tinggi diambil dari nilai yang tinggi pada Tabel 2 Berikut adalah daftar tabel centroid data yang dapat dilihat pada tabel 3 dibawah :

Tabel 3. Centroid Data Awal

C1	3025	18539	4804	1164	6204	51676	12729	2483	9253	96256	46244	4166
C2	685	3523	981	337	1215	11128	3395	788	1485	21058	10786	1341
C3	265	1254	354	90	776	3806	1212	132	1802	6186	3513	191

4. Menghitung Jarak Setiap Data Terhadap Centroid (Pusat Cluster)

Setelah data nilai pusat cluster awal ditentukan, maka langkah berikutnya yang akan dilakukan adalah menghitung jarak masing-masing data terhadap pusat cluster. Proses pencarian jarak terpendek pada Iterasi 1 dapat dilihat pada perhitungan dan tabel dibawah ini :

Hasil perhitungan jarak pusat Cluster dapat dilihat pada Tabel 4

Tabel 4. Hasil Perhitungan Jarak Pusat Cluster Iterasi 1

NO	KABUPATEN/KOTA	JUNI				JULI				AGUSTUS				C1	C2	C3	Jarak Terpendek	
		PDP	POSI TIF	SEM BUAH	MENINGGAL	PDP	POSI TIF	SEM BUAH	MENINGGAL	PDP	POSI TIF	SEM BUAH	MENINGGAL					
1	KAB. ASAHAN	54	262	64	37	76	894	244	149	8	1489	818	158	119541,373	24783,4657	6600,523464	6600,523464	
2	KAB. BATUBARA	0	54	0	0	47	861	287	0	114	1430	1087	0	119533,771	24771,8014	6565,972358	6565,972358	
3	KAB. DAIRI	23	48	30	0	10	143	41	0	97	451	164	42	120991,923	26241,4486	8003,394155	8003,394155	
4	KAB. DELI DERDANG	685	3523	981	337	1215	11128	3395	788	1485	21058	10786	1341	94792,4762	0	18446,54523	0	
5	KAB. HUMBANG HASUNDUTAN	12	15	0	0	19	80	39	0	32	221	163	0	121213,923	26460,2629	8219,192783	8219,192783	
6	KAB. KARO	43	327	69	0	67	1117	397	76	787	2572	965	235	118448,711	23717,0728	5475,779579	5475,779579	
7	KAB. LABUHANBATU	11	32	30	0	20	127	55	2	50	391	183	31	121043,684	26290,8077	8055,993483	8055,993483	
8	KAB. LABUHANBATU SELATAN	2	8	0	0	21	79	0	0	0	160	86	0	121299,704	26547,5151	8309,194486	8309,194486	
9	KAB. LABUHANBATU UTARA	28	60	0	60	0	62	0	62	113	203	19	66	121278,254	26529,5328	8280,95695	8280,95695	
10	KAB. LANGKAT	13	223	48	63	187	831	167	92	730	1596	717	142	119479,606	24747,4923	6445,874805	6445,874805	
11	KAB. MANDAILING NATAL	25	45	22	0	38	60	60	0	8	104	68	0	121344,163	26591,7263	8347,032946	8347,032946	
12	KAB. NIAS	0	0	0	0	0	0	0	0	0	18	0	0	121480,877	26728,92	8486,425337	8486,425337	
13	KAB. NIAS BARAT	0	0	0	0	0	0	0	0	18	2	0	0	121492,183	26740,5229	8494,262711	8494,262711	
14	KAB. NIAS SELATAN	0	0	0	0	5	23	0	0	47	37	18	2	121445,295	26694,228	8444,467242	8444,467242	
15	KAB. NIAS UTARA	2	0	0	0	0	3	0	0	0	48	16	0	121449,695	26697,5518	8456,606412	8456,606412	
16	KAB. PADANG LAWAS	5	34	0	8	0	84	23	31	0	126	33	37	121337,336	26585,2777	8346,995987	8346,995987	
17	KAB. PADANG LAWAS UTARA	57	22	3	0	15	46	31	0	4	92	55	0	121372,551	26620,0819	8377,659578	8377,659578	
18	KAB. PAKPAK BHARAT	0	0	0	0	0	0	0	0	0	2	0	0	121493,553	26741,5164	8498,061367	8498,061367	
19	KAB. SAMOSIR	0	8	0	0	25	60	35	2	26	88	71	0	121364,697	26612,0948	8365,316969	8365,316969	
20	KAB. SERDANG BEDAGAI	62	394	142	30	81	906	414	31	160	2745	1150	71	118368,17	23617,7598	5523,060474	5523,060474	
21	KAB. SIMALUNGUN	509	1325	369	31	728	3140	1309	128	1314	4388	2997	237	115038,23	20346,2072	2064,676246	2064,676246	
22	KAB. TAPANULI SELATAN	0	31	13	0	14	93	41	0	19	187	157	2	121235,918	26482,3361	8240,847044	8240,847044	
23	KAB. TAPANULI TENGAH	5	13	0	0	26	164	14	39	22	1105	751	62	120255,747	25496,5341	7324,060349	7324,060349	
24	KAB. TAPANULI UTARA	47	158	92	0	31	260	164	0	35	441	301	0	120870,358	26117,1675	7877,251551	7877,251551	
25	KAB. TOBA	31	60	60	0	38	290	125	19	47	578	365	31	120742,625	25987,022	7755,086009	7755,086009	
26	KOTA PEMATANGSIANTAR	265	1254	354	90	776	3806	1212	132	1802	6186	3513	191	113121,062	18446,5452	0	0	
27	KOTA BINJAI	43	407	79	27	72	1291	508	96	111	2861	1459	266	117981,731	23218,2087	5128,712022	5128,712022	
28	KOTA GUNUNGSITOLI	0	12	0	0	0	34	21	0	6	240	31	18	121273,676	26522,1997	8291,384384	8291,384384	
29	KOTA MEDAN	3025	18539	4804	1164	6204	51676	12729	2483	9253	96256	46244	4166	0	94792,4762	113121,062	0	0
30	KOTA PADANG SIDEMPUAN	0	74	0	30	19	168	71	41	0	297	178	80	121097,099	26342,448	8109,985635	8109,985635	
31	KOTA SIBOLGA	5	5	0	0	12	263	0	8	316	513	208	49	120870,627	26126,0683	7855,965631	7855,965631	
32	KOTA TANJUNG BALAI	13	77	0	49	0	190	2	62	0	310	109	62	121111,142	26359,9374	8130,245937	8130,245937	
33	KOTA TEBING TINGGI	27	110	30	0	38	528	108	28	513	1550	654	99	119719,751	24978,6636	6733,08087	6733,08087	

5. Menentukan Posisi Cluster atau Pengelompokan

Dalam menentukan posisi cluster masing-masing data hasil banyaknya Penyebaran Covid-19 di Sumatera Utara berdasarkan jarak minimum data terhadap pusat cluster.





Tabel 7. Hasil Perhitungan Jarak Pusat Cluster Iterasi 2

NO	KABUPATEN/KOTA	JUNI				JULI				AGUSTUS				C1	C2	C3	Jarak Terpendek
		PDP	POSI TIF	SEM BUH	MENI NGGAL	PDP	POSI TIF	SEM BUH	MENI NGGAL	PDP	POSI TIF	SEM BUH	MENI NGGAL				
1	KAB. ASAHAN	54	262	64	37	76	894	244	149	8	1489	818	158	119541,373	24783,4657	756,839547	756,839547
2	KAB. BATUBARA	0	54	0	0	47	861	287	0	114	1430	1087	0	119533,771	24771,8014	827,8995595	827,8995595
3	KAB. DAIRI	23	48	30	0	10	143	41	0	97	451	164	42	120991,923	26241,4486	769,5856677	769,5856677
4	KAB. DELI DERDANG	685	3523	981	337	1215	11128	3395	788	1485	21058	10786	1341	94792,4762	0	25483,29126	0
5	KAB. HUMBANG HASUNDUTAN	12	15	0	0	19	80	39	0	32	221	163	0	121213,923	26460,2629	985,65117	985,65117
6	KAB. KARO	43	327	69	0	67	1117	397	76	787	2572	965	235	118448,711	23717,0728	1882,980625	1882,980625
7	KAB. LABUHANBATU	11	32	30	0	20	127	55	2	50	391	183	31	121043,684	26290,8077	818,8591611	818,8591611
8	KAB. LABUHANBATU SELATAN	2	8	0	0	21	79	0	0	0	160	86	0	121299,704	26547,5151	1075,905172	1075,905172
9	KAB. LABUHANBATU UTARA	28	60	0	60	0	62	0	62	113	203	19	66	121278,254	26529,5328	1058,031908	1058,031908
10	KAB. LANGKAT	13	223	48	63	187	831	167	92	730	1596	717	142	119479,606	24747,4923	908,3146232	908,3146232
11	KAB. MANDAILING NATAL	25	45	22	0	38	60	60	0	8	104	68	0	121344,163	26591,7263	1118,263488	1118,263488
12	KAB. NIAS	0	0	0	0	0	0	0	0	0	18	0	0	121480,877	26728,92	1254,22517	1254,22517
13	KAB. NIAS BARAT	0	0	0	0	0	0	0	0	18	2	0	0	121492,183	26740,5229	1263,760085	1263,760085
14	KAB. NIAS SELATAN	0	0	0	0	5	23	0	0	47	37	18	2	121445,295	26694,228	1215,500801	1215,500801
15	KAB. NIAS UTARA	2	0	0	0	0	3	0	0	0	48	16	0	121449,695	26697,5518	1223,267242	1223,267242
16	KAB. PADANG LAWAS	5	34	0	8	0	84	23	31	0	126	33	37	121337,336	26585,2777	1115,046608	1115,046608
17	KAB. PADANG LAWAS UTARA	57	22	3	0	15	46	31	0	4	92	55	0	121372,551	26620,0819	1146,696899	1146,696899
18	KAB. PAKPAK BHARAT	0	0	0	0	0	0	0	0	0	2	0	0	121493,553	26741,5164	1266,55968	1266,55968
19	KAB. SAMOSIR	0	8	0	0	25	60	35	2	26	88	71	0	121364,697	26612,0948	1135,650433	1135,650433
20	KAB. SERDANG BEDAGAI	62	394	142	30	81	906	414	31	160	2745	1150	71	118368,17	23617,7598	1945,068621	1945,068621
21	KAB. SIMALUNGUN	509	1325	369	31	728	3140	1309	128	1314	4388	2997	237	115038,23	20346,2072	5414,175442	5414,175442
22	KAB. TAPANULI SELATAN	0	31	13	0	14	93	41	0	19	187	157	2	121235,918	26482,3361	1008,871633	1008,871633
23	KAB. TAPANULI TENGAH	5	13	0	0	26	164	14	39	22	1105	751	62	120255,747	25496,5341	518,0167234	518,0167234
24	KAB. TAPANULI UTARA	47	158	92	0	31	260	164	0	35	441	301	0	120870,358	26117,1675	664,0834843	664,0834843
25	KAB. TOBA	31	60	60	0	38	290	125	19	47	578	365	31	120742,625	25987,022	525,2352498	525,2352498
26	KOTA PEMATANGSIANTAR	265	1254	354	90	776	3806	1212	132	1802	6186	3513	191	113121,062	18446,5452	7238,375916	7238,375916
27	KOTA BINJAI	43	407	79	27	72	1291	508	96	111	2861	1459	266	117981,731	23218,2087	2291,115759	2291,115759
28	KOTA GUNUNGSITOLI	0	12	0	0	0	34	21	0	6	240	31	18	121273,676	26522,1997	1055,854189	1055,854189
29	KOTA MEDAN	3025	18539	4804	1164	6204	51676	12729	2483	9253	96256	46244	4166	0	94792,4762	120234,5405	0
30	KOTA PADANG SIDEMPUAN	0	74	0	30	19	168	71	41	0	297	178	80	121097,099	26342,448	878,3229639	878,3229639
31	KOTA SIBOLGA	5	5	0	0	12	263	0	8	316	513	208	49	120870,627	26126,0683	674,3084653	674,3084653
32	KOTA TANJUNG BALAI	13	77	0	49	0	190	2	62	0	310	109	62	121111,142	26359,9374	902,1375708	902,1375708
33	KOTA TEBING TINGGI	27	110	30	0	38	528	108	28	513	1550	654	99	119719,751	24978,6636	666,9440441	666,9440441

Tabel 8. Hasil Cluster Iterasi 2

NO	NAMA KABUPATEN/KOTA	KELOMPOK
1	KAB. ASAHAN	3
2	KAB. BATUBARA	3
3	KAB. DAIRI	3
4	KAB. DELI DERDANG	2
5	KAB. HUMBANG HASUNDUTAN	3
6	KAB. KARO	3
7	KAB. LABUHANBATU	3
8	KAB. LABUHANBATU SELATAN	3
9	KAB. LABUHANBATU UTARA	3
10	KAB. LANGKAT	3
11	KAB. MANDAILING NATAL	3
12	KAB. NIAS	3
13	KAB. NIAS BARAT	3
14	KAB. NIAS SELATAN	3



---

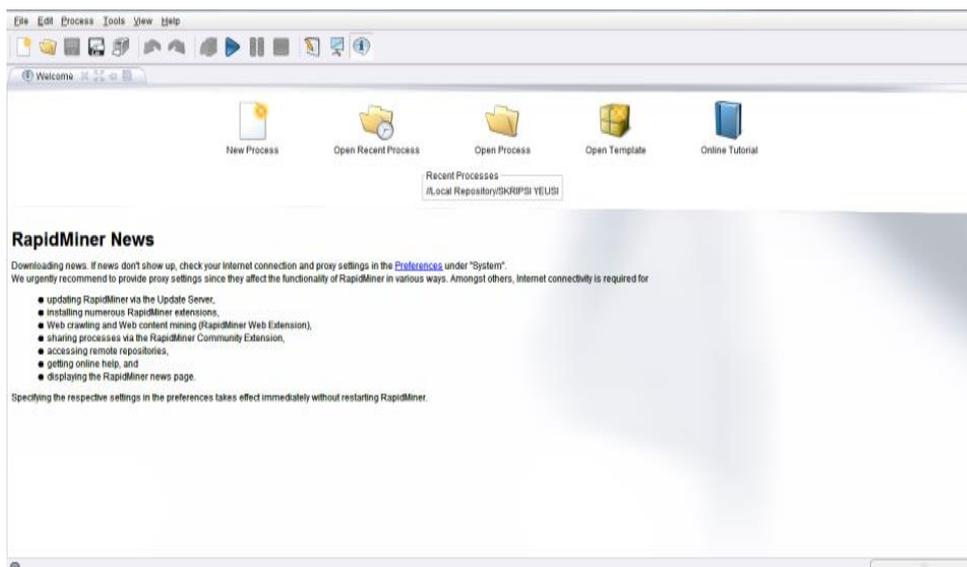
	1
	1
	1
	1
1	
	1
	1
	1
	1

---

Perhitungan manual pada data diatas didapatkan hasil akhir dimana pada iterasi 1 dan iterasi 2 pengelompokan data yang dilakukan terhadap 3 cluster didapatkan hasil yang sama. Hasil dari kedua iterasi tersebut berniali  $C1=1$ ,  $C2=1$ , dan  $C3=31$  pada posisi data tiap cluster. Sehingga posisi cluster pada data tersebut tidak mengalami perubahan maka proses iterasi berhenti sampai iterasi 2.

#### 4.1. Tampilan Menu Utama RapidMiner

Berikut akan dijelaskan bagian dari menu awal pada tools RapidMiner seperti tampilan awal dari RapidMiner dan akan dimulai dengan menu New Process. Bisa dilihat pada gambar 4.



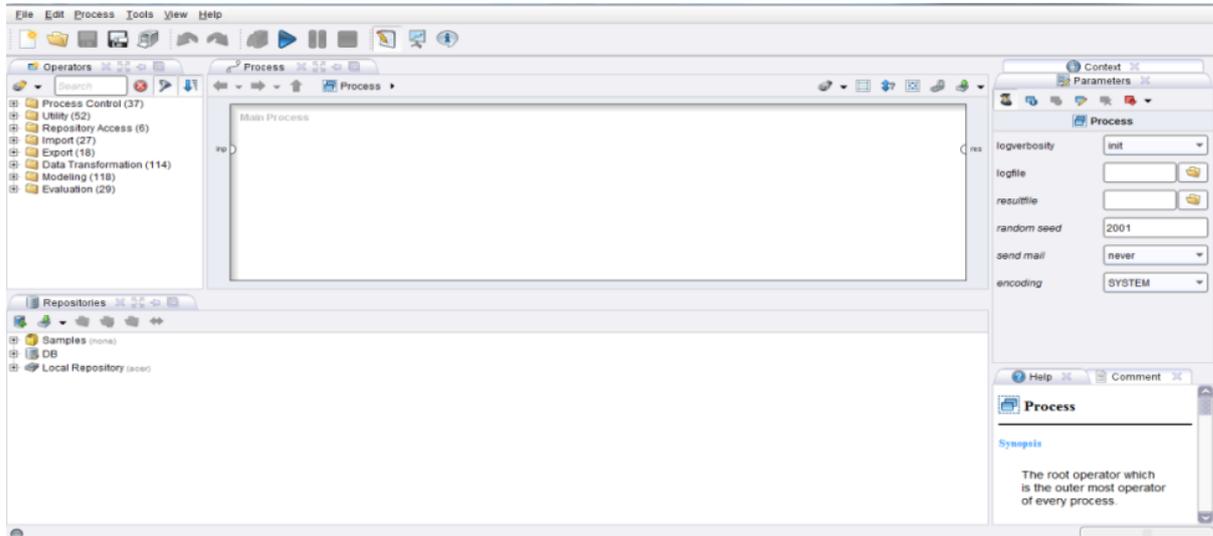
Gambar 4. Tampilan Awal RapidMiner

Pada tampilan berikut terdapat beberapa menu seperti New Process, Open Recent Process, Open Process, Open Template dan Online Tutorial. Penjelasan dari menu tersebut adalah sebagai berikut :

1. New Process, menu ini berisikan tampilan awal jika akan membuat proses baru dengan data baru yang belum pernah ada sebelumnya dalam tools.
2. Open Recent Process, menu ini digunakan untuk menampilkan atau membuka proses yang baru saja dibuat dengan jangka waktu tertentu dan biasanya akan langsung muncul pada tampilan awal tools tanpa harus mencari pada penyimpanan lokal.
3. Open Process, menu ini berfungsi untuk membuka proses yang telah dibuat sebelumnya dan tersimpan di penyimpanan lokal.
4. Open Template, menu ini berisi pilihan-pilihan yang telah disediakan oleh tools RapidMiner.
5. Open Tutorial, menu ini digunakan untuk memulai tutorial/ tahapan secara online. Menu ini juga berfungsi untuk memberikan pengenalan dan beberapa konsep data mining.

#### 4.2. New Process

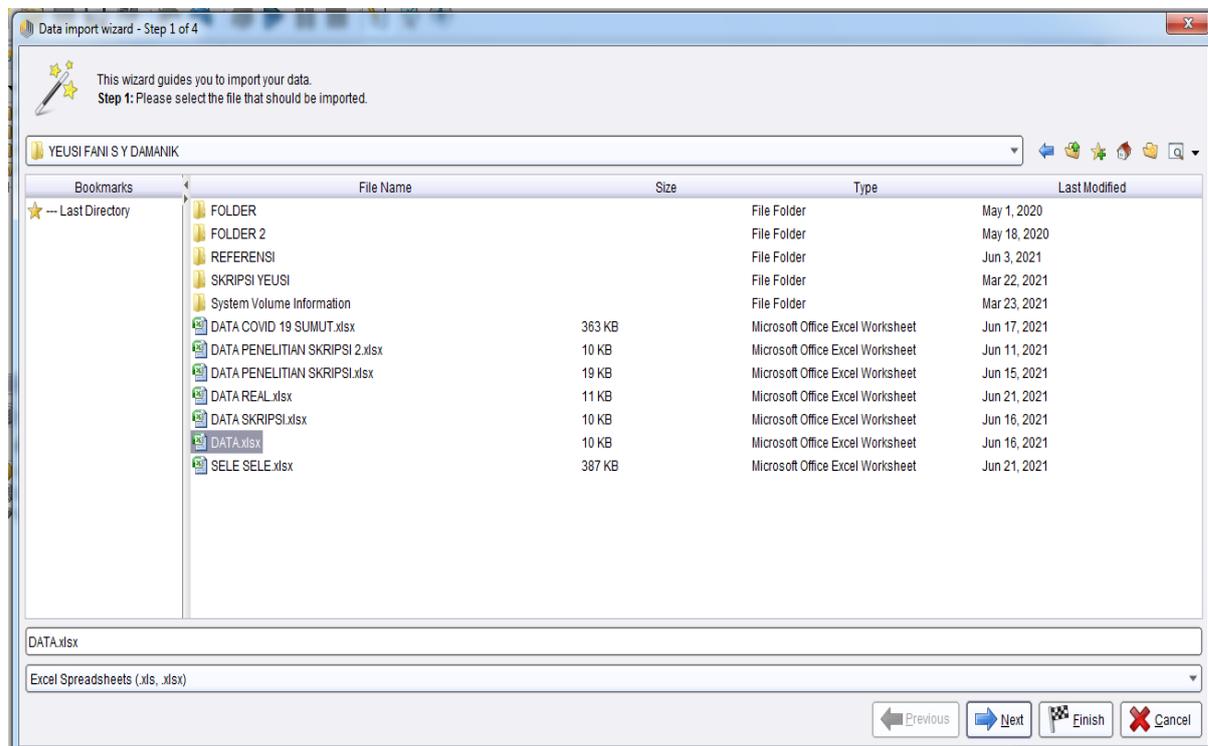
Karena data yang akan diuji adalah data baru, maka pilih New Process. Maka akan muncul jendela seperti gambar 5.



Gambar 5. Tampilan New Process

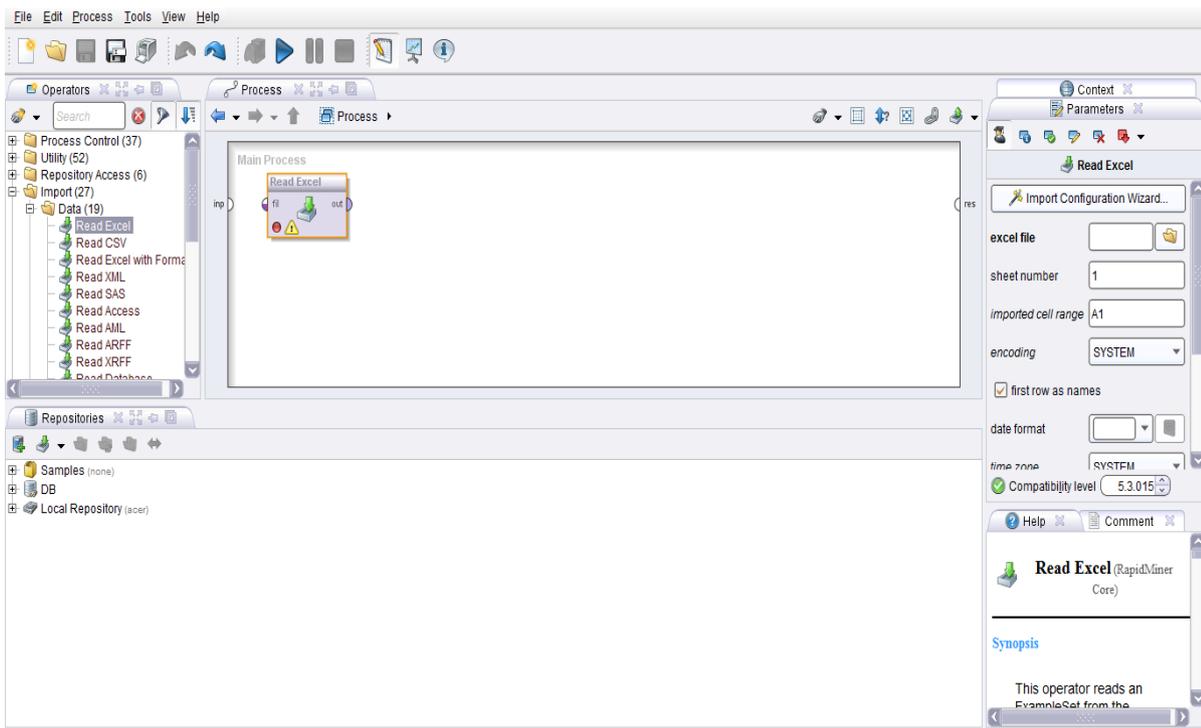
#### 4.3. Input Sistem

Sistem menjelaskan cara memasukkan data baru yang akan dieksekusi lebih lanjut, pada tahap ini data yang akan di eksekusi berupa data Excel. Tahapan dapat dilihat pada gambar 6



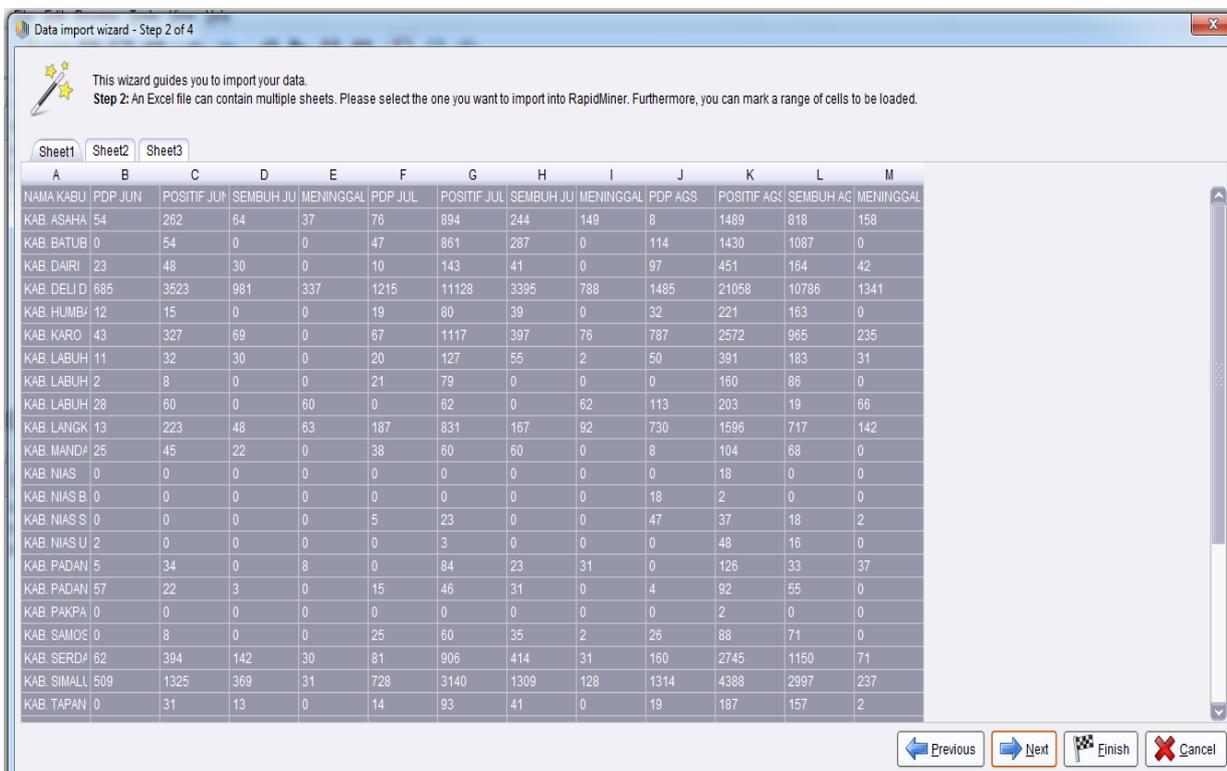
Gambar 6. Data Import Tahap 1

Untuk memasukkan data Excel yang berisi data penelitian dapat dilakukan dengan cara klik Import Configuration Wizard lalu klik file yang akan di import seperti tampilan gambar 7



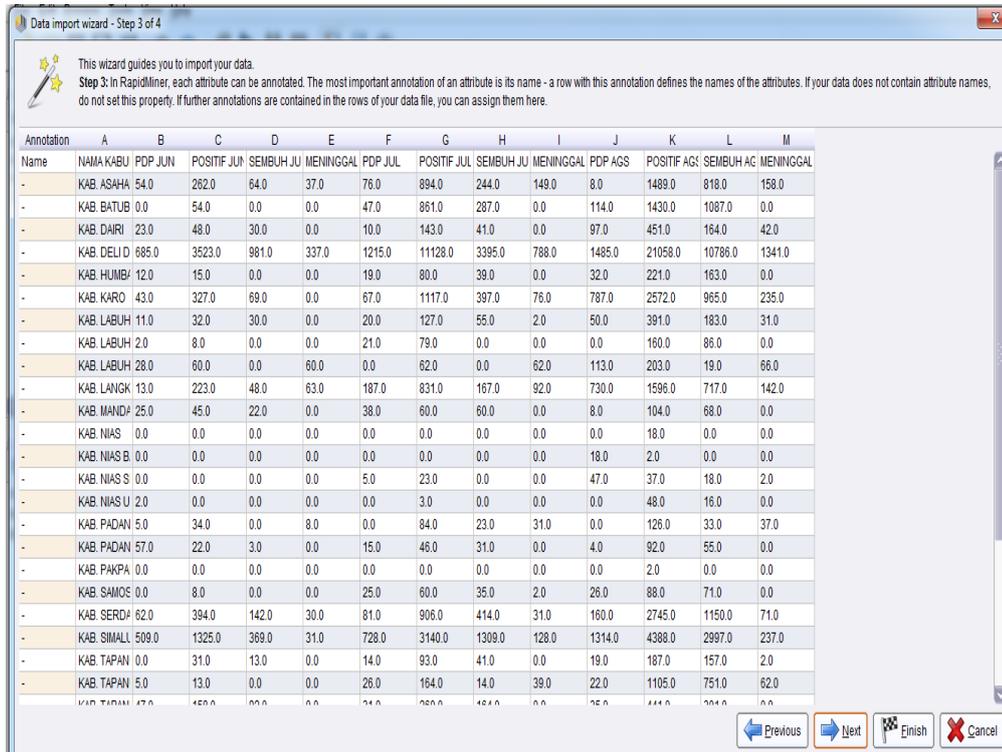
Gambar 7. Tampilan Import Data Excel

Setelah menentukan file yang akan di import klik Next maka akan muncul tampilan seperti gambar 8 yang menunjukkan isi file yang berisi data yang akan diproses.



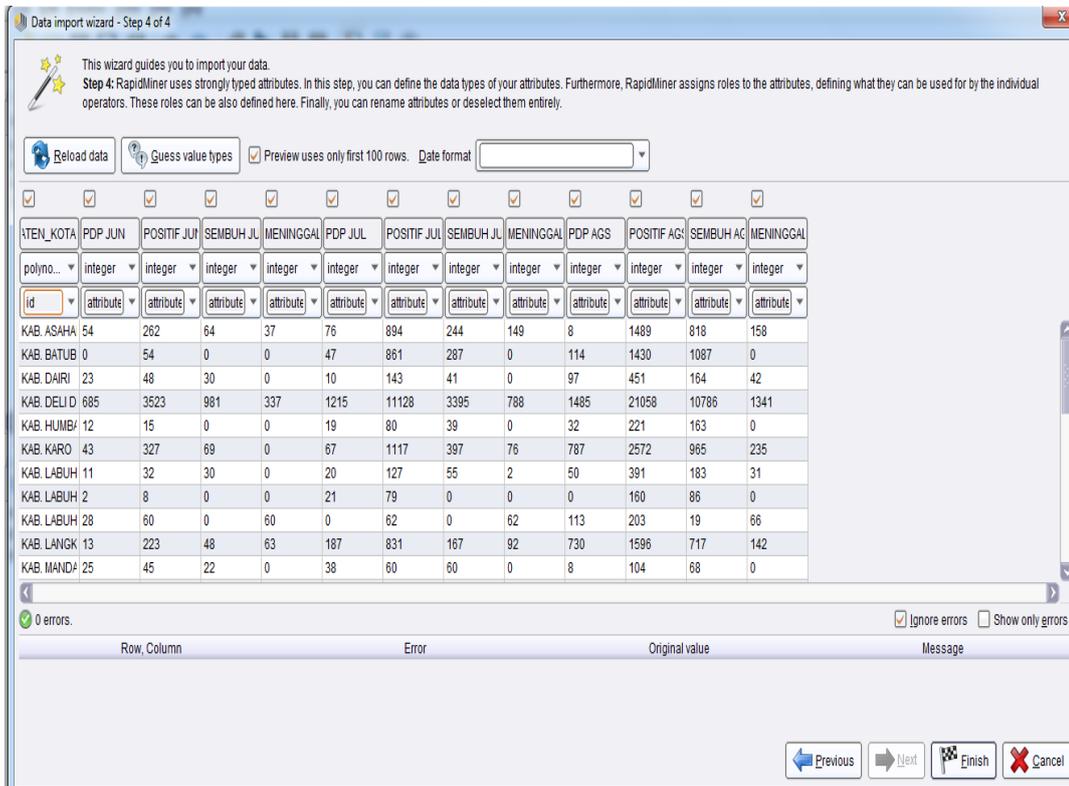
Gambar 8. Tampilan Import Data Cluster Tahap 2

Selanjutnya klik *Next* maka akan muncul tampilan seperti gambar 9.



Gambar 9. Tampilan Import Data Cluster Tahap 3.

Lalu klik Next dan pilih tempat penyimpanan data kemudian tentukan nama data yang akan diproses. Selanjutnya klik Finish. Data yang telah di import akan tampak seperti gambar 10. dan data siap untuk diproses

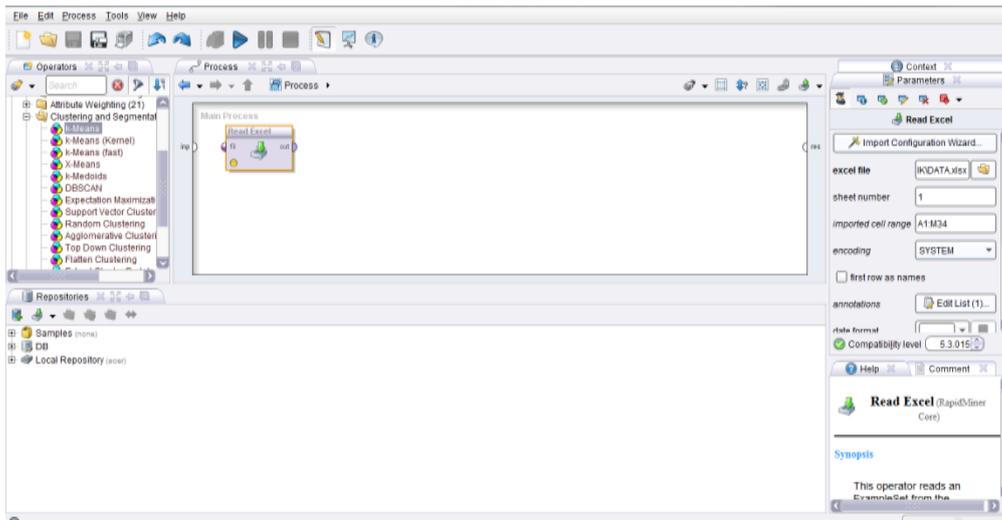


Gambar 10. Tampilan Import Data Cluster Tahap 4

Pada tahap 4 dilakukan pemilihan tipe data dimana bagian “No attribute” diubah menjadi tipe data ”Id”, lalu klik *Next*

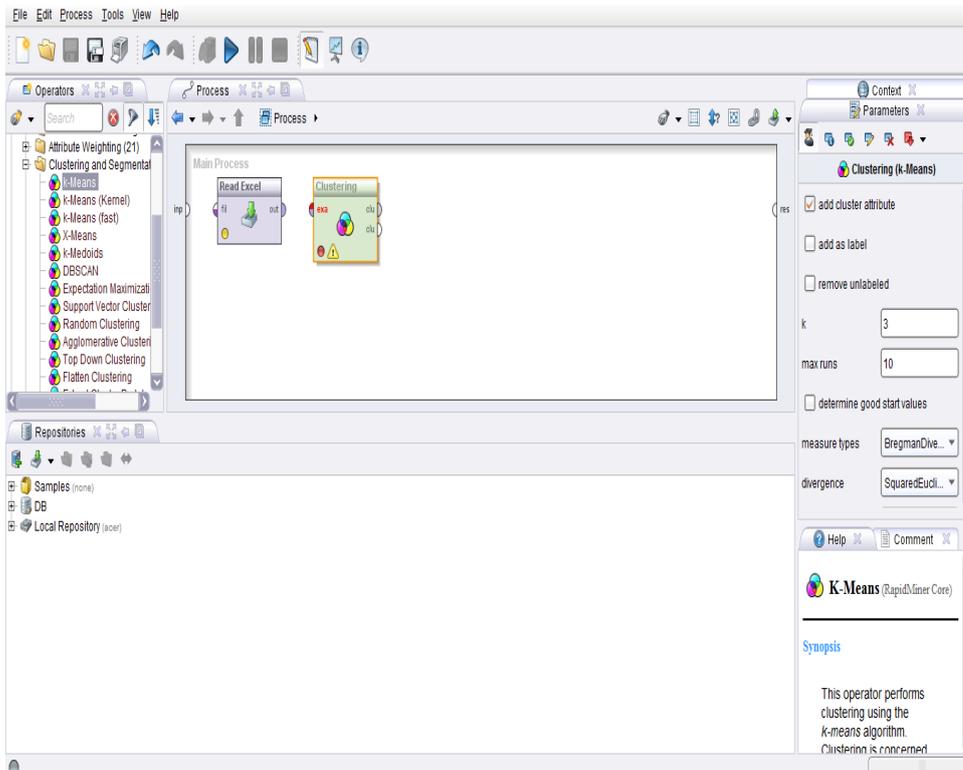
#### 4.4. Pemrosesan Sistem

Pada tahap ini akan dijelaskan tahapan-tahapan proses penggunaan K-Means didalam RapidMiner data yang telah di import. Tahapan pertama yaitu dengan mengimport file excel yang akan diproses seperti tampilan gambar 11. berikut:



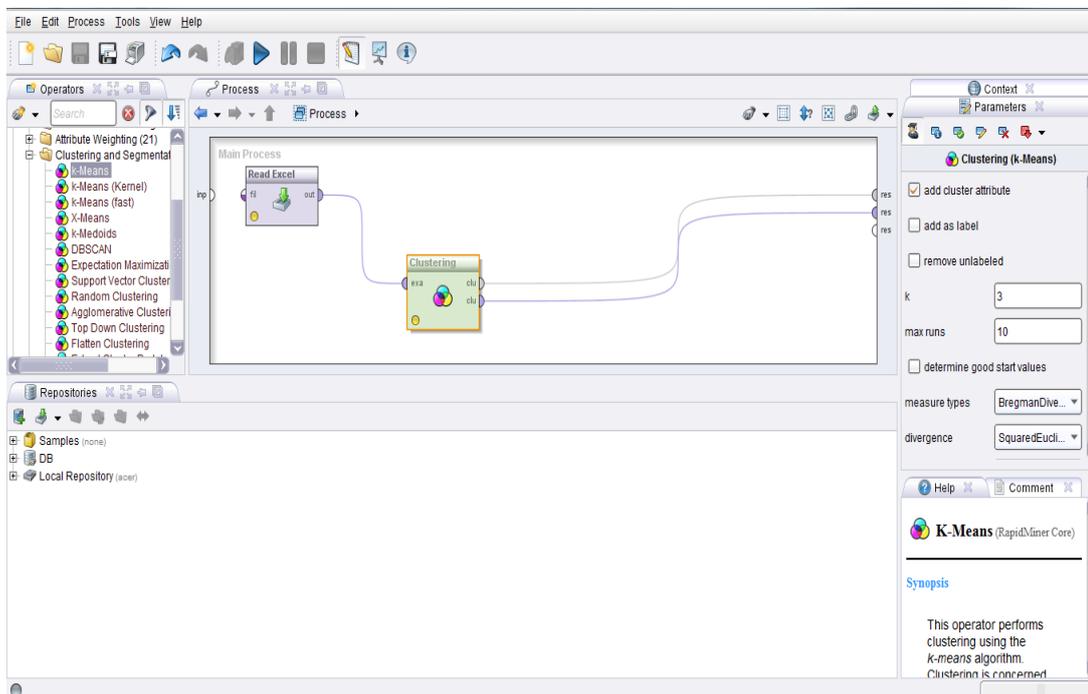
Gambar 11. Pemrosesan Tahap 1

Kemudian meng klik Clustering and Segmentation lalu pilih K-Means untuk membentuk pengelompokan, data perlu dihubungkan dengan operator Clustering. Dapat dilihat pada gambar 12



Gambar 12. Pemrosesan Tahap 2

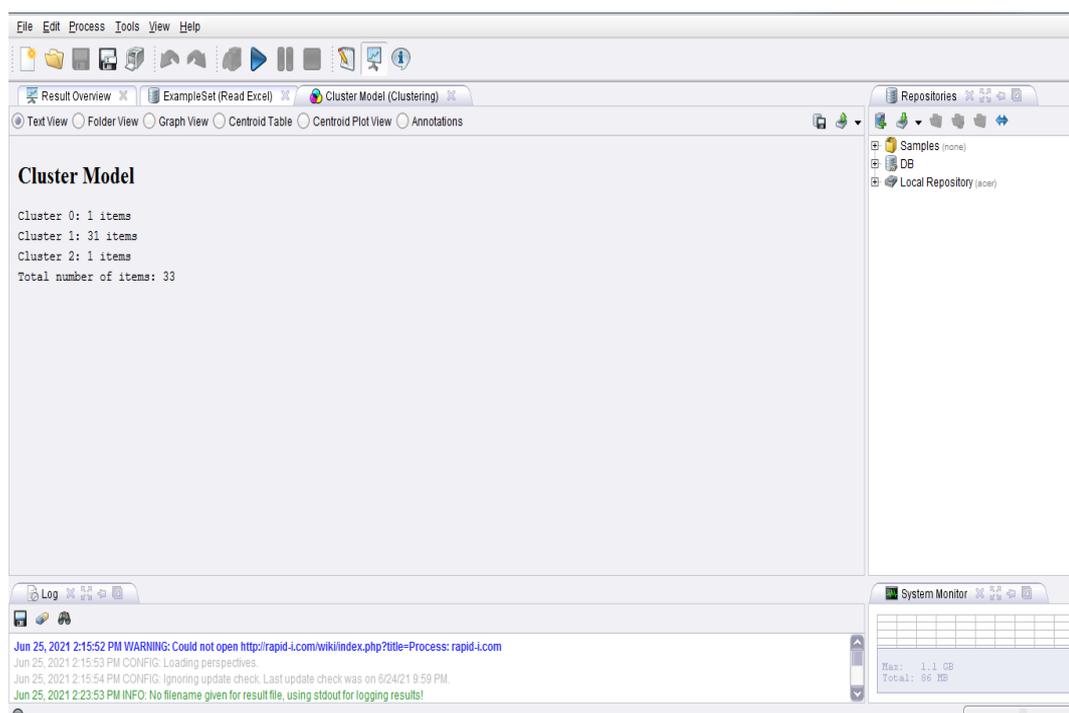
Kemudian hubungkan antar Read Excel dengan Clustering seperti gambar dibawah ini:



Gambar 13. Pemrosesan Tahap 3

#### 4.5. Output Sistem RapidMiner

Untuk mendapatkan hasil pengelompokan maka pada tahap selanjutnya dapat dilakukan dengan cara mengklik tanda panah biru yang terdapat dibagian tengah atas atau pada bagian Toolbar. Pada tahap ini akan menampilkan hasil akhir serta langkah terakhir dalam penggunaan tools RapidMiner. Dapat dilihat pada gambar 14.



Gambar 14. Nilai Cluster Model RapidMiner

Keterangan :

- a) Jumlah Cluster 0 (tinggi) berjumlah 1 items
- b) Jumlah Cluster 2 (sedang) berjumlah 1 items
- c) Jumlah Cluster 1 (rendah) berjumlah 31 items

Sehingga dapat diketahui hasil pengelompokan dari RapidMiner berikut ini dapat dilihat pada gambar 15.



Gambar 15. Hasil Pengelompokan *RapidMiner*

Berdasarkan gambar 15. dapat diketahui bahwa pada kelompok tinggi memiliki sebanyak 1 item, kelompok sedang memiliki sebanyak 1 item, sedangkan kelompok rendah memiliki sebanyak 31 item.

#### 4.6. Pembahasan

Berdasarkan penjelasan di atas mengenai tahap-tahap penggunaan serta hasil yang telah ditampilkan maka berikut akan membahas mengenai keterkaitan dari hasil yang didapat antara perhitungan manual algoritma dengan hasil yang ditampilkan oleh tools RapidMiner dan dapat disimpulkan bahwa hasil perhitungan manual dari algoritma K-Means dan dari aplikasi yaitu Software Microsoft Office Excel 2007 hasilnya sama dan dapat di input ke dalam aplikasi RapidMiner.

Row No.	id	cluster	PDP JUN	POSITIF JUN	SEMBUH J.	MENINGGA.	PDP JUL	POSITIF JUL	SEMBUH J.	MENINGGA.	PDP AGS	POSITIF AGS	SEMBUH A.	MENINGGA.
1	1	cluster_1	54	262	64	37	76	894	244	149	8	1489	818	158
2	2	cluster_1	0	54	0	0	47	861	287	0	114	1430	1087	0
3	3	cluster_1	23	48	30	0	10	143	41	0	97	451	164	42
4	4	cluster_2	685	3523	981	337	1215	11128	3395	788	1485	21058	10786	1341
5	5	cluster_1	12	15	0	0	19	80	39	0	32	221	163	0
6	6	cluster_1	43	327	69	0	67	1117	397	76	787	2572	965	235
7	7	cluster_1	11	32	30	0	20	127	55	2	50	391	183	31
8	8	cluster_1	2	8	0	0	21	79	0	0	0	160	86	0
9	9	cluster_1	28	60	0	60	0	62	0	62	113	203	19	66
10	10	cluster_1	13	223	48	63	187	831	167	92	730	1596	717	142
11	11	cluster_1	25	45	22	0	38	60	60	0	8	104	68	0
12	12	cluster_1	0	0	0	0	0	0	0	0	0	18	0	0
13	13	cluster_1	0	0	0	0	0	0	0	0	18	2	0	0
14	14	cluster_1	0	0	0	0	5	23	0	0	47	37	18	2
15	15	cluster_1	2	0	0	0	0	3	0	0	0	48	16	0
16	16	cluster_1	5	34	0	8	0	84	23	31	0	126	33	37
17	17	cluster_1	57	22	3	0	15	46	31	0	4	92	55	0
18	18	cluster_1	0	0	0	0	0	0	0	0	0	2	0	0
19	19	cluster_1	0	8	0	0	25	60	35	2	26	88	71	0

Gambar 16. Tampilan Data Perhitungan RapidMiner

## 5. KESIMPULAN DAN SARAN

Berdasarkan pembahasan dan hasil implementasi software RapidMiner pada pengelompokan penyebaran Covid-19 di Sumatera Utara, maka dapat ditarik kesimpulannya yaitu data mining dengan algoritma K-Means dapat diterapkan untuk mengelompokkan banyaknya penyebaran Covid-19 di Sumatera Utara. Sumber data yang digunakan pada penelitian ini adalah data yang diperoleh dari Media Center Gugus Tugas Covid-19 SUMUT. Jumlah data yang di uji sebanyak 33 Kota/Kabupaten di Sumatera Utara dengan menggunakan tiga cluster, yaitu untuk cluster tinggi berjumlah 1 data yaitu Kota Medan, untuk cluster sedang berjumlah 1 data yaitu Kab. Deli Serdang, untuk cluster rendah berjumlah 31 data yaitu Kab. Asahan, Kab. Batubara, Kab. Dairi, Kab. Humbang Hasundutan, Kab. Karo, Kab. Labuhanbatu, Kab. Labuhanbatu Selatan, Kab. Labuhanbatu Utara, Kab. Langkat, Kab. Mandailing Natal, Kab. Nias, Kab. Nias Barat, Kab. Nias Selatan, Kab. Nias Utara, Kab. Padang Lawas, Kab. Padang Lawas Utara, Kab. Pakpak Bharat, Kab. Samosir, Kab. Serdang Bedagai, Kab. Simalungun, Kab. Tapanuli Selatan, Kab. Tapanuli Tengah, Kab. Tapanuli Utara, Kab. Toba, Kota Pematangsiantar, Kota Binjai, Kota Gunungsitoli, Kota Padang sidempuan, Kota Sibolga, Kota Tanjung Balai dan Kota Tebing Tinggi dan hasil yang diperoleh dari algoritma K-Means dapat diterapkan ke dalam RapidMiner dengan nilai validasi yang sama dan dapat menjadi masukan kepada pemerintah Sumatera Utara agar dapat memberikan penanganan dan bantuan untuk daerah yang terjangkit Covid-19 dan mengedukasi masyarakat untuk menghindari daerah yang tingkat penyebaran Covid-19nya tinggi sehingga dapat memutus mata rantai penyebaran Covid-19 di Sumatera Utara.

Sedangkan saran dari penelitian ini dapat dijadikan rekomendasi terhadap penelitian selanjutnya guna menjadikan hasil dari penelitian yang lebih baik lagi untuk kedepannya sehingga bisa diterima oleh masyarakat, menjadi masukan bagi pemerintah khususnya di Sumatera Utara serta penelitian ini juga dapat digunakan sebagai referensi dan dapat dikembangkan untuk penulisan artikel ilmiah.

## DAFTAR PUSTAKA

- [1] A. Susilo *et al.*, "Coronavirus Disease 2019: Tinjauan Literatur Terkini," *J. Penyakit Dalam Indones.*, vol. 7, no. 1, p. 45, 2020, doi: 10.7454/jpdi.v7i1.415.
- [2] A. R. Setiawan and S. Ilmiyah, "Lembar Kegiatan Siswa untuk Pembelajaran Jarak Jauh Berdasarkan Literasi Saintifik pada Topik Penyakit Coronavirus 2019 (COVID-19)," vol. 2, no. 1, pp. 28–36, 2020, doi: 10.35542/osf.io/h4632.
- [3] N. Jannah and T. Yulianto, "Mengelompokkan Siswa Berprestasi Akademik dengan Menggunakan Metode K Means Kelas VII MT," *Zeta - Math J.*, vol. 2, no. 2, pp. 41–45, 2016.
- [4] J. S. Gulo and S. R. Andani, "K-Means Dalam Memprediksi Siswa Yang Layak Menerima Bantuan Siswa Miskin Pada Sd Negeri 127696 Pematangsiantar," *Pros. SiManTap Semin. Nas. Mat. dan Terap.*, vol. 1, pp. 829–837, 2019.
- [5] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, p. 167, 2018, doi: 10.14421/jiska.2018.23-06.
- [6] D. Setiyadi and A. L. I. Nurdin, "Data mining potensi akademik siswa berbasis online 1 1,2," *J. Ilm. SISFOTENIKA*, vol. 2, no. 19, pp. 31–40, 2014.
- [7] Maria Goreti usboko, "No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title," *Gastrointest. Endosc.*, vol. 10, no. 1, pp. 279–288, 2018, [Online]. Available: <http://dx.doi.org/10.1053/j.gastro.2014.05.023><https://doi.org/10.1016/j.gie.2018.04.013><http://www.ncbi.nlm.nih.gov/pubmed/29451164><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5838726><http://dx.doi.org/10.1016/j.gie.2013.07.022>.
- [8] E. Fammaldo and L. Hakim, "Penerapan Algoritma K-Means Clustering Untuk

- Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar,” *J. Ilm. Teknol. Infomasi Terap.*, vol. 5, no. 1, pp. 23–31, 2019, doi: 10.33197/jitter.vol5.iss1.2018.249.
- [9] F. E. M. Agustin, “Implementasi Algoritma K-Means Untuk Menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional (Studi Kasus: Smp Negeri 101 Jakarta),” *J. Tek. Inform.*, vol. 8, no. 1, pp. 73–78, 2015, doi: 10.15408/jti.v8i1.1938.
- [10] V. No and N. Mona, “Konsep Isolasi Dalam Jaringan Sosial Untuk Meminimalisasi Efek Contagious (Kasus Penyebaran Virus Corona Di Indonesia),” *J. Sos. Hum. Terap.*, vol. 2, no. 2, pp. 117–125, 2020, doi: 10.7454/jsht.v2i2.86.
- [11] K. Fatmawati and A. P. Windarto, “Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi,” *Comput. Eng. Sci. Syst. J.*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.
- [12] P. Hariyati, S. Saifullah, and M. Fauzan, “Tehnik Data Mining Dalam Mengelompokkan Kasus Pneumonia Pada Balita Berdasarkan Provinsi Di Indonesia,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, 2019, doi: 10.30865/komik.v3i1.1670.
- [13] Z. Zahrotunnimah, “Langkah Taktis Pemerintah Daerah Dalam Pencegahan Penyebaran Virus Corona Covid-19 di Indonesia,” *SALAM J. Sos. dan Budaya Syar-i*, vol. 7, no. 3, pp. 247–260, 2020, doi: 10.15408/sjsbs.v7i3.15103.