

Clustering Data Pasien Covid Berdasarkan Usia dan Gejala Menggunakan Algoritma K-Means

Nurhidayati^{1*}, Lola Mauliya², Suhartini³

^{1,2}Program Studi Sistem Informasi, Universitas Hamzanwadi

³Program Studi Teknik Informatika, Universitas Hamzanwadi
hidayati2188@gmail.com

Abstrak

Kasus COVID-19 sudah tidak asing lagi terdengar ditelinga setiap masyarakat, namun masih banyak sekali masyarakat yang tidak peduli akan pentingnya menjaga protocol kesehatan karena kurangnya edukasi mengenai virus COVID-19. Peningkatan kasus covid-19 semakin tinggi dari tahun ketahun dengan gejala pasien yang semakin beragam, oleh karena itu peneliti tertarik melakukan penelitian tentang pengelompokan pasien covid berdasarkan usia dan gejala menggunakan metode clustering algoritma k-means. Rapidminer 5 merupakan aplikasi yang digunakan untuk mengolah data pasien covid-19, namun sebelum melakukan pengolahan data akan dilakukan beberapa proses dan tahapan diantaranya yaitu cleaning data, seleksi data, intergrasi, transformation sampai data siap diolah menggunakan aplikasi rapidmider. Dari hasil yang didapatkan pada cluster 1 yang berjumlah 866 items yakni terdiri dari pasien dengan rentang usia 42-81 tahun dengan jumlah kasus kematian 50 orang, sedangkan pada cluster ke 2 berjumlah 1566 items dengan rentang usia 1-41 tahun dengan kasus kematian 16 orang. Metode clustering algoritma k-means menjadi solusi yang paling tepat untuk mengolah data pasien covid-19, ini karena tingkat akurasi yang dihasilkan cukup efektif. Hasil yang didapatkan dalam pengelompokan data pasien covid-19 menggunakan algoritma k-means bisa dijadikan sebagai bahan evaluasi oleh pemerintah khususnya Dinas Kesehatan Kabupaten Lombok Timur dalam mengatasi kasus penyebaran covid-19 yang semakin tinggi.

Kata kunci : *Data Mining, Clustering, K-Means, Rapidminer, Data covid-19*

Abstract

The case of COVID-19 is familiar to everyone's ears, but there are still a lot of people who don't care about the importance of maintaining health protocols because of the lack of education about the COVID-19 virus. The increase in cases of covid-19 is getting higher from year to year with increasingly diverse patient symptoms, therefore researchers are interested in conducting research on grouping covid patients based on age and symptoms using the k-means algorithm clustering method. Rapidminer 5 is an application used to process patient data. covid-19, but before processing the data, several processes and stages will be carried out including data cleaning, data selection, integration, transformation until the data is ready to be processed using the rapidmider application. From the result obtained cluster 1 wich amounted to 866 items it was certain that it consisted of patients aged 42-81 years with 50 deaths, while in the cluster 2 wich amounted to 1566 items with an age range of 1-41 years with 16 death. The k-means algorithm clustering method is the most appropriate solution for processing covid patient data. -19, this is because the level of accuracy produced is quite effective. The results obtained in grouping Covid-19 patient data using the k-means algorithm can be used as evaluation material by the government, especially the East Lombok District Health Office in dealing with cases of the increasingly high spread of COVID-19

Keywords : *Data Mining, Clustering, K-Means, Rapidminer, Covid-19 data*

1. Pendahuluan

Penyakit menular adalah penyakit yang disebabkan karena virus, parasit, jamur, dan bakteri. Banyak masyarakat yang khawatir terhadap penyakit menular namun seperti yang diketahui bahwa penyakit menular ini dapat dicegah dengan cara vaksinasi dan kesadaran pola hidup masyarakat yang sehat dan bersih. Pada akhir tahun 2019 lalu telah ditemukan penyakit menular terbaru yang disebut *Coronavirus Disease-19 (COVID-19)*, virus ini ditemukan di kota Wuhan, China dan sekarang penyakit menular ini sudah tersebar dengan cepat hampir diseluruh negara bahkan di Indonesia. Saat ini belum ditemukan pasti penyebab virus corona, namun diketahui bahwa virus ini disebarkan oleh hewan. Virus ini juga mampu ditularkan dari satu spesies ke spesies lainnya[1]. Hal ini membuat beberapa Negara memberlakukan kebijakan yang disebut *lockdown*. Bahkan di Indonesia juga memberlakukan kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) demi mencegah percepatan penularan virus corona (Covid-19).

Virus corona (Covid-19) adalah kumpulan virus yang menginfeksi gangguan pernapasan, pernapasan ringan seperti flu. Namun, virus ini juga bisa menyebabkan infeksi pernapasan berat, seperti infeksi paru-paru. Selain virus SARS-CoV-2 atau virus corona, virus yang termasuk dalam

kelompok Coronavirus adalah virus *Severe Acute Respiratory Syndrom (SARS)* dan virus penyebab *Middle East Respiratory Syndrome (MERS)*. Menurut data yang dirilis hingga 15 Maret 2022 tercatat 5.914.532 orang yang terkonfirmasi positif covid-19 yang mengakibatkan jumlah kematian 152.745 jiwa[2]. Sedangkan update terakhir pada tanggal 17 Agustus jumlah kasus kematian di NTB 998 jiwa. Berdasarkan usia, rata-rata usia di atas (> 60 tahun) memiliki persentase angka kematian yang lebih tinggi dibandingkan kelompok usia lainnya. Dengan demikian peran pemerintah sangatlah penting dalam mengambil kebijakan dalam membatasi pergerakan masyarakat dan penyebaran virus yang waktu itu dikenal dengan istilah *Social and physical distancing*, atau PSBB dan yang terbaru adalah aturan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM).

Penelitian ini dilakukan pada Dinas Kesehatan Kabupaten Lombok Timur yang dimana data yang akan diolah adalah data private. Data private yaitu data mentah yang belum dipublish yang didapatkan dari tempat penelitian. Diketahui bahwa pada lokasi penelitian yakni Dinas Kesehatan Kabupaten Lombok Timur masih melakukan pengolahan data secara manual yakni menggunakan excel dimana dengan pengolahan seperti itu membutuhkan waktu yang lama dalam proses penginputan data, dan data tersebut hanya berisi data pasien terkonfirmasi positif

covid-19 dan disini peneliti ingin mengolah data tersebut untuk mengelompokkan jumlah pasien yang terkonfirmasi positif covid-19 sesuai dengan usia dan gejalanya. Pada penelitian ini penulis ingin menerapkan data mining menggunakan algoritma K-Means Clustering. Data mining adalah proses menemukan pola dan pengetahuan yang menarik dari suatu data dalam jumlah yang besar. Data mining menggambarkan sebuah pengumpulan teknik-teknik dengan tujuan untuk menemukan pola-pola yang tidak diketahui pada data yang telah dikumpulkan. Algoritma K-Means merupakan salah satu algoritma Clustering yang masuk ke dalam kelompok Unsupervised learning yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok dengan sistem partisi.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

- Fadlahu Warsita (2021). Dalam penelitiannya yang berjudul "Penerapan Metode Clustering Dengan Algoritma K-Means Untuk Mengelompokkan Data Gizi Bayi Pada Pelayanan Posyandu Desa Perian Kecamatan Montong Gading". Peneliti melakukan pengolahan data dimana data yang di ambil adalah data private, kemudian nantinya data tersebut akan dikelompokkan sesuai karakteristik gizi masing-masing anak (analisis hasil pengklasteran) melalui dua

parameter yaitu berat badan (BB), tinggi badan (TB) dan berat badan menurut umur (BB/U)[3].

- Nurma Yulia Rusmin Saragih, Indah Ambarita (2021). Dalam jurnal yang berjudul "Data Mining Pengelompokkan Anak Stunting Berdasarkan Usia, Penyebab Dab Pekerjaan Orang Tua Dengan Menggunakan Metode Clustering (Studi Kasus : Dinas Kesehatan Kabupaten Langkat)". Dalam penelitian ini peneliti melakukan pengelompokkan data anak untuk mengantisipasi bertambahnya kasus kurang gizi pada balita, serta pentingnya partisipasi dari orang tua dan petugas pos pelayanan terpadu, puskesmas atau posyandu untuk memantau nilai gizi pada balita yang sangat dibutuhkan. Diketahui hasil dari 1000 data yang telah digunakan untuk centroid 1 (2.03 2.42 2.65) dikelompokkan sebagai data anak stunting pada grup usia 13-20 bulan dengan penyebab yaitu tidak mendapatkan asi eksklusif dan pekerjaan orang tuanya ialah wiraswasta, serta centroid 2 (4.05 4.40 3.43) dikelompokkan data stunting pada grup usia adalah 29-36 bulan dengan penyebab yaitu janin kurang asupan makanan dan pekerjaan orang tuanya adalah buruh, centroid 3 (4.91 2.07 2.63) dikelompokkan sebagai data anak stunting pada grup usia 37-44 bulan dengan penyebab yaitu tidak mendapatkan asi

- eksklusif dan pekerjaan orang tua wiraswasta[4].
- Achmad Solichin, Khansa Khairunnisa (2020). Dalam jurnal yang berjudul "Clusterisasi Persebaran Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means". Dalam penelitian ini peneliti melakukan pengelompokan data persebaran virus corona menggunakan algoritma k-means clustering. Tujuan penelitian ini digunakan untuk mengcluster persebaran virus corona berdasarkan parameter jumlah ODP, PDP, kasus positif, pasien sembuh dan data meninggal. Penelitian ini menggunakan 3 cluster, dihasilkan pengelompokan C1 sebanyak 19 kecamatan, C2 sebanyak 23 kecamatan dan C3 sebanyak 2 kecamatan[5].
 - Muhammad Saiful, Syamsudin (2021). Dalam jurnal yang berjudul "Implementasi Algoritma Naïve Bayes Untuk Memprediksi Predikat Ketuntasan Belajar Siswa Pasca Pandemi Covid-19". Dalam penelitian ini peneliti menyimpulkan bahwa dalam menentukan data tuntas dan tidak tuntas dapat di prediksi dan dievaluasi dengan memanfaatkan teknik data mining menggunakan algoritma naïve bayes. Algoritma ini mampu menganalisa pola data tuntas dan tidak tuntas, dalam mengambil kebijakan untuk melengkapi data siswa yang dikategorikan tuntas dan tidak tuntas dengan memanfaatkan teknik data mining. Setelah dilakukan nya pengujian sebanyak 9 kali yaitu dari validation 2 sampai dengan validation 10 maka didapatkan dua hasil akurasi terbaik dari K-Fold Validation 4 dan K-Fold Validation 5 yang masing-masing memiliki nilai accuracy sebesar 83,89% pada K-Fold Validation 4 dan 82,74% dari K-Fold Validation 5. Dan selisih akurasi keduanya adalah 0.01% yang menunjukkan bahwa terjadi peningkatan 0,80% pada rentang K-Fold Validation 4 dan 8. Peningkatan pada setiap uji coba tidak selalu sama ini dipengaruhi oleh seberapa banyak pembagian data yang dilakukan[6].
 - Muhammad Saiful, LM Samsu, Fathurrahman (2021). Dalam jurnal yang berjudul "Sistem Deteksi Infeksi COVID-19 Pada Hasil X-Ray Rontgen Menggunakan Algoritma Convolutional Neural Network (CNN). Dari hasil analisis dan pembahasan dalam studi kasus ini, peneliti membuat kesimpulan dengan menggunakan 25.000 step dan 1.000 evaluation_step untuk melakukan pengujian model4 pada data baru terinfeksi Covid-19 menghasilkan akurasi sebesar 98%. Tingkat akurasi pendeteksian thorax terinfeksi covid-19 pada model4 menggunakan algoritma CNN dapat dinilai bekerja dengan baik dengan beragam tingkat akurasi pada hasil pengujian berkisar 60-99%[7].

2.2. Landasan Teori

1. Sejarah Dan Definisi Corona Virus Disease 19 (COVID-19)

Diakhir tahun 2019 virus Corona untuk pertama kali ditemukan di kota Wuhan, China. Virus ini mulai masuk ke Indonesia dan menyebar pada bulan Maret 2020 hingga menggemparkan dunia, karena virus ini sejumlah Negara langsung menerapkan lockdown untuk memutus rantai penyebaran Covid-19.

Covid merupakan penyakit menular, dan ditandai dengan gejala pada bagian pernapasan akut (*Coronavirus 2 (severe acute respiratory syndrome coronavirus 2 atau SARS-CoV-2)*). Virus ini adalah bagian dari keluarga besar Coronavirus yang dapat menyerang manusia dan hewan. Penularannya dapat terjadi apabila adanya kontak antar sesama manusia. Ketika menyerang manusia, penyakit yang disebabkan oleh virus ini akan menyerang saluran pernapasan yang mengakibatkan infeksi saluran pernapasan, flu, MERD (*Middle East Respiratory Syndrom*)[8].

2. Data Mining

Data mining merupakan bagian dari konsep teknologi informasi yang berkaitan dengan data dan informasi dan data mining juga dapat digunakan sebagai metode atau teknik pemodelan untuk menganalisis data dengan variasi data yang banyak untuk mendapatkan pola dan hubungan antar variasi tersebut.[9]

Data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Keluaran data mining dapat digunakan untuk membantu pengambilan keputusan dimasa depan.[10]

Istilah data mining memiliki beberapa pandangan, seperti *knowledge discover* atau *pattern recognition*. Kedua istilah tersebut sebenarnya memiliki ketetapan masing-masing, istilah *knowledge discovery* atau penemuan pengetahuan tepat karena digunakan tujuan dari data mining memang untuk mendapat pengetahuan yang masih tersembunyi didalam bongkahan data.[11]

3. Clustering

Cluster atau analisis cluster adalah proses pembentukan kelompok data (cluster) dari himpunan data yang tidak diketahui kelompok-kelompok atau kelas-kelasnya dan proses menentukan data-data termasuk dalam cluster yang mana. Cluster merupakan proses untuk mengetahui kelas-kelas taksonomi atau batryologi, atau analisis topologi data data-data yang ada[12].

Analisis Cluster adalah bagaimana bisa mengetahui pola pembelian barang oleh konsumen pada waktu-waktu tertentu. Dengan mengetahui pola kelompok pembelian tersebut, maka perusahaan/retailer dapat menentukan jadwal promosi yang dapat diberikan sehingga

dapat membantu meningkatkan omzet penjualan[13].

4. K-Means

K-Means merupakan salah satu metode pengelompokan data non-hierarki yang mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan kedalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain. Adapun tujuan pengelompokan data ini adalah untuk meminimalkan fungsi objek yang di set dalam satu kelompok dan memaksimalkan variasi antar kelompok[14].

5. Rapid Miner

Rapid Miner adalah perangkat lunak yang bersifat terbuka. Rapid miner digunakan untuk melakukan analisis prediksi dan memiliki lebih dari 500 operator data mining, termasuk operator input, output, data preprocessing dan visualisasi. Rapid miner menggunakan bahasa java jadi dapat dimengerti oleh semua system informasi. Rapid miner menggunakan GUI (*Graphic User Interface*) untuk merancang sebuah pipeline analitis. GUI akan menghasilkan XML (*Extensible Mark Up Language*) yang mendefinisikan proses analisis keinginan pengguna untuk diterapkan ke data[15]

3. Metode Penelitian

Adapun beberapa metode penelitian yang dilakukan untuk memperoleh informasi yang dibutuhkan adalah :

1. Observasi Lapangan

Pengumpulan data yang dilakukan dengan pengamatan secara langsung di lokasi penelitian yakni pada dinas kesehatan Lombok Timur, guna untuk mendapatkan informasi terkait data covid-19 dan untuk mengetahui bagaimana permasalahan yang ada disana.

2. Studi Pustaka

Pengumpulan informasi yang dilakukan dengan cara mencari informasi dan referensi yang berhubungan dengan penelitian, referensi diperoleh dari buku, jurnal / artikel dan internet.

3. Wawancara

Tanya jawab langsung dengan pihak terkait yakni salah satu pegawai Dinas Kesehatan Lombok Timur untuk mendapatkan informasi yang dibutuhkan peneliti agar lebih memahami permasalahan yang terjadi.

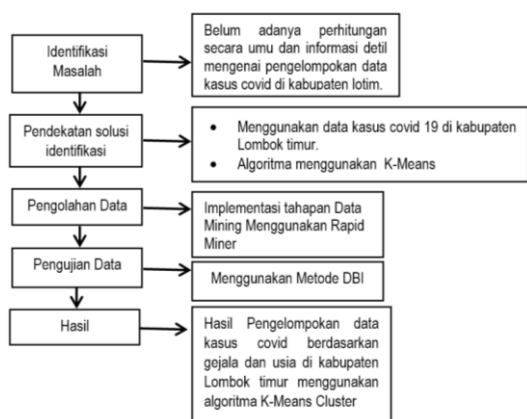
3.2 Lokasi Penelitian

Dalam penelitian “clustering data pasien covid berdasarkan usia dan gejala menggunakan algoritma k-means”, peneliti mengambil lokasi pada Dinas Kesehatan Kabupaten Lombok Timur.

Adapun jenis datasetnya adalah termasuk dataset privat karna data tersebut tidak di publish.

3.3 Tahapan Penelitian

Tahapan penelitian merupakan langkah-langkah yang dilakukan oleh peneliti, berikut adalah tahapan yang dilakukan oleh peneliti dapat dilihat pada Gambar 1 dibawah ini:



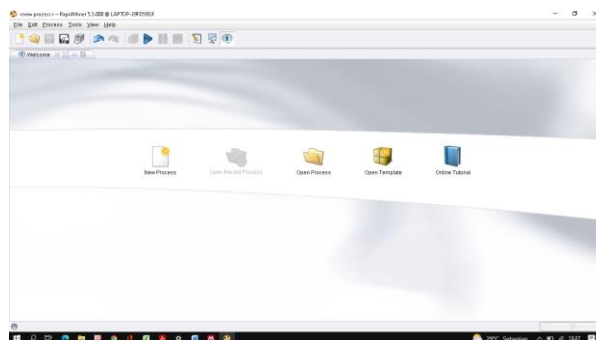
Gambar 1. Tahapan Penelitian

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

Pada penelitian ini, peneliti menggunakan Software Rapid Miner sebagai alat untuk melakukan pengolahan data Covid-19. Berikut ini adalah tahapan pengolahan data Covid-19 menggunakan algoritma K-Means untuk mengelompokkan data pasien Covid berdasarkan usia dan gejala :

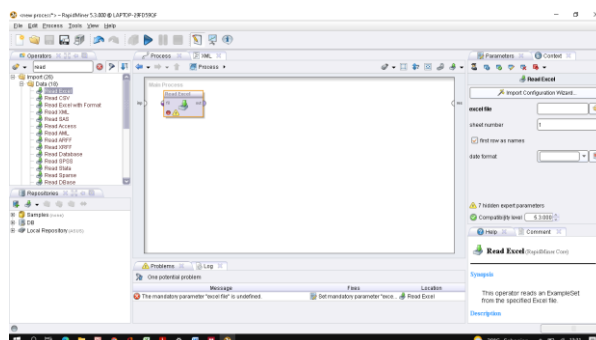
1. Tampilan Utama



Gambar 2. Tampilan Utama

Halaman registrasi digunakan untuk melakukan pendaftaran sebagai users jika customers belum memiliki akun.

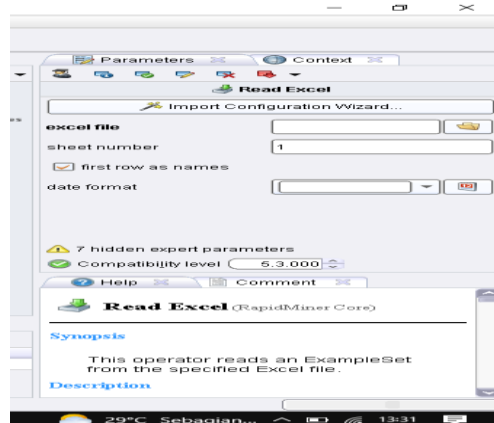
2. Lembar kerja baru (tampilan read Excel)



Gambar 3. Lembar kerja baru (tampilan read Excel)

Setelah masuk lembar kerja baru kemudian klik read excel pada operators kemudian tarik read excel ke lembar kerja seperti gambar diatas.

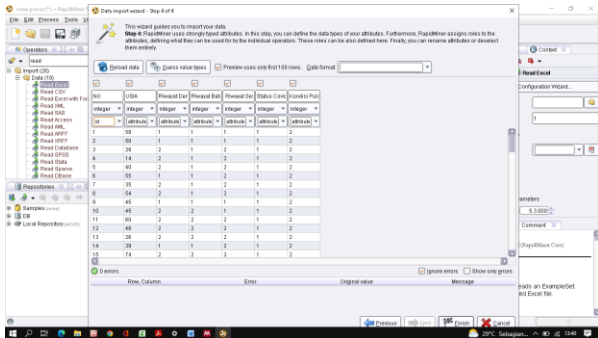
3. Pengimputan Data



Gambar 4 Pengimputan Data

Setelah memasukkan read excel lembar kerja, selanjutnya akan dilakukan pengimputan data atau import configuration wizard.

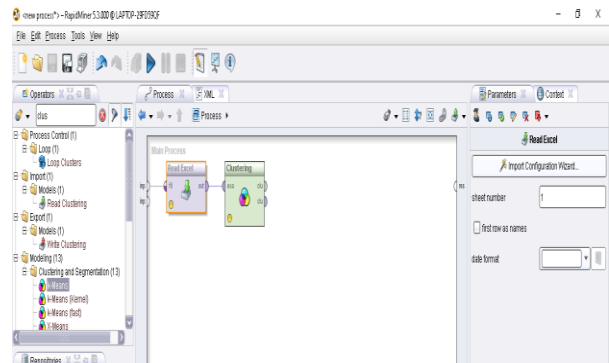
4. Pemilihan Atribut



Gambar 5. Pemilihan Atribut

Setelah sebelumnya mengklik Import Coviguration Wizard selanjutnya kita akan diarahkan untuk mencari data, selanjutnya pilih data set yang akan di olah pada aplikasi rapid miner, pastikan data yang akan dimasukkan sudah di cleaning sebelumnya, jika datanya sudah siap diinput maka klik next sampai tampilan rapid miner berbentuk seperti gambar diatas, setelah tampilan sudah seperti atas maka akan dilakukan pemilihan atribut. Pada analisis data Covid ini, bagian atribut nomor di ubah menjadi id, selanjutnya atribut gejala sampai dengan atribut kondisi pasien tidak perlu dirubah. Jika sudah selesai menentukan atribut maka klik finish untuk lanjut ke tahap selanjutnya.

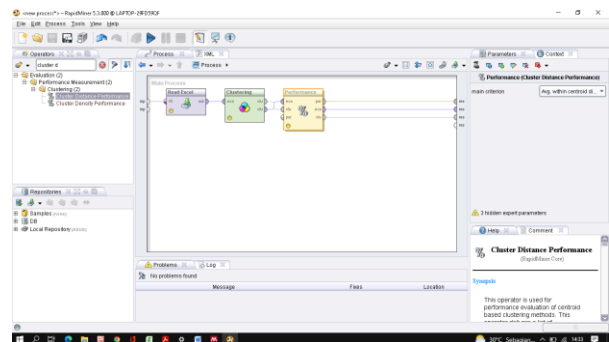
5. Implementasi Algoritma K-Means



Gambar 6. Implementasi Algoritma K-Means

Apabila pada tahap sebelumnya tidak terjadi error maka selanjutnya akan lanjut ketahap pengimplementasian algoritma K-Means seperti gambar diatas. Ketik K-Means pada tools operator, tarik K-means lembar kerja dan sambungkan dengan operator read excel.

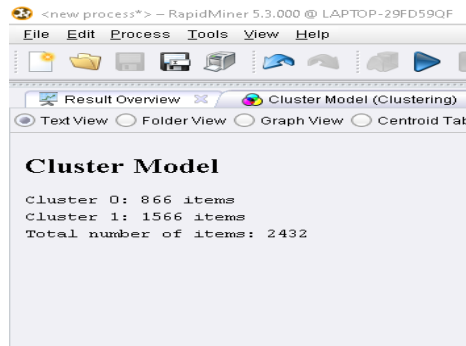
6. Operator Cluster Distance Performance



Gambar 7. Operator Cluster Distance Performance

Setelah selesai dengan tahap sebelumnya maka lanjut ketik cluster distance performance pada operator lalu tarik lembar kerja dan hubungkan setiap operator seperti gambar di atas. Langkah terakhir klik tools run. maka akan muncul hasil dari pengolahan data.

7. Cluster Model



Gambar 8. Cluster Model

Dari hasil pengolahan data diatas, terdapat 2 cluster. Pada cluster 0 berjumlah = 866 items dan cluster 1 = 1566.

8. Centroid

Attribute	cluster_0	cluster_1
USIA	55.258	27.375
Riwayat Den	1.539	1.771
Riwayat Bat	1.436	1.296
Riwayat Ses	1.498	1.509
Status Covid	1	1
Kondisi Puls	1.942	1.990

Gambar 9. Centroid

Centroid adalah titik pusat cluster atau jarak terdekat dalam menentukan setiap anggota kelompok cluster, dimana proses ini merupakan salah satu cara untuk menentukan jarak terdekat dari masing-masing cluster yang ada.

4.2 Pembahasan

Berdasarkan hasil analisis yang sudah didapatkan dari pengolahan data pasien Covid-19 menggunakan aplikasi *RapidMiner* dan perhitungan manual menggunakan excel bisa dikatakan sudah memiliki persentase nilai cluster

yang sama. Pada cluster 0 (cluster 1) berjumlah 866 items dengan rentang usia 42- 81 tahun dengan kasus meninggal 50 orang dan sehat 816 orang, sedangkan pada cluster 1 (Cluster 2) berjumlah 1566 item pada rentang usia 1-41 tahun dengan kasus meninggal lebih sedikit yaitu 16 orang dan pasien sehat berjumlah 1550 orang. Berikut ini adalah persentase nilai yang didapatkan dari perhitungan manual excel dan *Rapid Miner*

Tabel 1. Hasil Persentase Perhitungan Rapid MIner dan Excel

	Rapidminer	Excel	Persentase
C1	866	866	100%
C2	1566	1566	100%

5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan pada Dinas Kesehatan Kabupaten Lombok Timur, tentang pengolahan data pasien covid 19 menggunakan teknik analisis Data Mining dengan implementasi algoritma K-Means Cluster didapatkan dua kluster yaitu kluster usia dan gejala, yang merupakan tolak ukur pemerintah dalam memutus rantai penyebaran virus Covid-19 tahun 2020. Dari hasil pengolahan tersebut, didapatkan hasil pada kluster 1 (Berdasarkan usia) berjumlah 866 items yang terdiri dari pasien dengan rentang usia 42 – 81 tahun, dengan jumlah kasus kematian sebanyak 50 orang.

Sedangkan pada kluster 2 (Berdasarkan gejala) berjumlah 1566 items dengan rentang usia 1-41 tahun dengan kasus kematian 16 orang

6. Daftar Pustaka

- [1] A. Amalia and N. Sa'adah, "Dampak Wabah Covid-19 Terhadap Kegiatan Belajar Mengajar Di Indonesia," *J. Psikol.*, vol. 13, no. 2, pp. 214–225, 2020, doi: 10.35760/psi.2020.v13i2.3572.
- [2] D. Pittara, "COVID-19 adalah penyakit akibat infeksi virus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)." <https://www.alodokter.com/covid-19>
- [3] Fadlahu Warsita, "Penerapan Metode Clustering Dengan Algoritma K-Means Untuk Mengelompokkan Data Gizi Bayi Pada Pelayanan Posyandu Desa Perian Kecamatan Montong Gading," p. 2, 2021.
- [4] S. Kaputama *et al.*, "Data Mining Pengelompokan Anak Stunting Berdasarkan Usia , Penyebab dan Pekerjaan Orang Tua Dengan Menggunakan Metode Clustering (Studi Kasus: Dinas Kesehatan Kabupaten Langkat)," 2021.
- [5] M. M. K-means, A. Solichin, and K. Khairunnisa, "Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta," vol. 5, no. 2, 2020.
- [6] M. Saiful and S. Samsuddin, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Predikat Ketuntasan Belajar Siswa Pasca Pandemi Covid 19," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 29–38, 2021, doi: 10.29408/jit.v4i1.2982.
- [7] M. Saiful, L. M. Samsu, and F. Fathurrahman, "Sistem Deteksi Infeksi COVID-19 Pada Hasil X-Ray Rontgen menggunakan Algoritma Convolutional Neural Network (CNN)," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 2, pp. 217–227, 2021, doi: 10.29408/jit.v4i2.3582.
- [8] N. Dwitri, J. A. Tampubolon, S. Prayoga, F. I. R.H Zer, and D. Hartama, "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 128–132, 2020, doi: 10.36294/jurti.v4i1.1266.
- [9] J. A. Ginting, T. Informatika, and U. B. Mulia, "DATA MINING UNTUK ANALISA PENGAJUAN KREDIT DENGAN MENGGUNAKAN METODE LOGISTIK REGRESI Data Mining To Analisizes Loan Proposal Using Regresion Logistic Method," no. 2, pp. 164–169, 2019.
- [10] S. Informasi and S. Royal, "Seminar Nasional Royal (SENAR) 2018," vol. 9986, no. September, 2018.
- [11] A. Bastian *et al.*, "PENERAPAN ALGORITMA K-MEANS CLUSTERING ANALYSIS PADA PENYAKIT MENULAR MANUSIA (STUDI KASUS KABUPATEN MAJALENGKA)," no. 1, pp. 26–32, 2018.
- [12] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan," vol. 5, no. 1, pp. 62–66, 2019.
- [13] E. Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Jogjakarta: ANDI, 2014.
- [14] E. Prasetyo, *Data Mining : Konsep Dan Aplikasi Menggunakan MATLAB*. Yogyakarta: ANDI, 2012.
- [15] D. R. M.Makmum Effendi, "PREDIKSI PENJUALAN PRODUK ROTI MENGGUNAKAN ALGORITMA C4.5 PADA PT. PRIMA TOP BOGA," *SIGMA- J. Teknol. Pelita Bangsa*, vol. 9, no. September, 2018