

Penerapan Algoritma K-Means Clustering Dalam Proses Pengelompokan Kasus Meninggal Dunia Covid-19 Di Indonesia

Asep Saepudin^{*1}, Rizki Entis Sutisna², Christina Juliane³

Sekolah Tinggi Ilmu Manajemen Informatika dan Komputer LIKMI
e-mail: ¹azharsaepudin5@gmail.com, ²rizkientiss@gmail.com, ³christina.juliane@likmi.ac.id

Abstrak

Negara Indonesia masih bergelut dengan wabah covid-19 sampai saat ini, sama dengan negara lain di dunia. Jumlah kasus harian covid-19 di terus bertambah dengan seiring tingkat kesembuhan, akan tetapi tidak sedikit juga yang meninggal. Pemerintah Indonesia memberikan sosialisasi kepada masyarakat untuk melakukan physical distancing guna memutus rantai penyebaran covid-19 yang tersebar di berbagai wilayah Indonesia. Maka dari itu, pengumpulan data pasti banyak sekali, dari data yang banyak itu kita dapat melihat pola – pola menentukan pengelompokan penyebaran covid-19 yang dilakukan berdasarkan tes menggunakan algoritma k-means clustering. Hasil penelitian yaitu mengetahui daerah yang memiliki tingkat kasus meninggal dunia paling tinggi. Dan mengambil keputusan untuk daerah tingkat kasus meninggal tinggi maka termasuk kedalam zona merah dan harus segera diambil tindakan.

Kata kunci – Covid-19, Clustering, Data Mining

Abstract

The country of Indonesia is still struggling with the Covid-19 outbreak to date, the same as other countries in the world. The number of daily cases of covid-19 continues to grow along with the rate of cure, but not a few also die. The Indonesian government provides socialization to the public to carry out physical distancing to break the chain of the spread of COVID-19 which is spreading in various parts of Indonesia. Therefore, there must be a lot of data collection, from that much data we can see patterns in determining the grouping of the spread of Covid-19 based on tests using the k-means clustering algorithm. The result of the research is to know the area that has the highest death rate. And making decisions for areas with a high death rate will be included in the red zone and action must be taken immediately.

Keywords - Covid-19, Clustering, Data Mining

1. PENDAHULUAN

Coronavirus adalah kelompok virus yang dapat menyebabkan penyakit pada hewan atau manusia[1]. Beberapa jenis *coronavirus* diketahui menyebabkan infeksi saluran pernafasan pada manusia mulai dari batuk, flu hingga yang lebih serius seperti *Middle East Respiratory Syndrome (MERS)* dan *Severe Acute Respiratory Syndrome (SARS)*. Dengan bertambahnya kasus kematian yang disebabkan covid-19, maka perlu adanya tindakan untuk



mencegah kematian. Pencegahan dapat dimulai dari diri sendiri seperti menggunakan masker dan patuh terhadap protokol kesehatan[2]. Tidak cukup dengan menjalankan prosedur kesehatan untuk mencegah penularan dan angka kematian, melainkan bisa menggunakan teknologi agar dapat mengetahui daerah yang memiliki angka kematian tinggi. Sehingga dapat dilakukan tindakan atau pengetatan dan pembatasan kegiatan masyarakat sesuai dengan aturan yang ditetapkan pemerintah.

Penggunaan teknologi data mining metode clustering dengan menggunakan algoritma K-Means, maka data tersebut dapat dikelompokkan berdasarkan angka kematian. Sehingga teknologi data mining dapat membantu berkontribusi dalam hal mengolah data. Data mining adalah kegiatan pengumpulan data atau menambang data, kemudian data tersebut diolah untuk menemukan pengetahuan, informasi, pola atau hubungan data yang berukuran sangat besar dan tidak dapat diukur [3,4]. Data *mining* tidak berdiri sendiri karena berkaitan dengan bidang ilmu lainnya seperti *database*, statistik, dan *artificial intelligent*.

1. *Database* – merupakan media penyimpanan data yang digunakan dalam *data mining* yang bersumber dari *database* tempat dimana data itu disimpan dan data tersebut digali atau dicari informasinya sehingga nanti akan digunakan untuk di olah ke proses selanjutnya.
2. Statistik – sebelum pengambilan keputusan, penggunaan statistik diperlukan dalam pengumpulan data, serta pengambilan sampel data dan probabilitas. Dalam dunia *data mining* penentuan sample data, cara menganalisa, dan mempresentasikan *output* yang digunakan menggunakan teknik statistik.
3. *Artificial intelligent* – adalah sebuah kecerdasan buatan yang dapat membantu dalam proses data mining.

Menurut Sani susanto, Dedy suryadi(2010),Universitas katolik parahyangan, berjudul “Pengantar data mining” dalam bukunya menjelaskan data mining adalah salah satu tujuan untuk mencari pengetahuan yang berada didalam data yang sangat banyak[5].

Menurut khasanah Irene Radius Sarreta (2021), Cermati.com, berjudul “Data mining: pengertian, fungsi, penerapan dan metode pengambilannya” data mining adalah proses mencari sebuah data yang diinginkan dari sekumpulan data yang sangat banyak untuk mendapatkan informasi yang dapat digunakan dan bermanfaat[6].

Dalam melakukan suatu penelitian diperlukan dukungan dari hasil penelitian terdahulu yang telah ada sebelumnya dan berkaitan dengan penelitian ini diantaranya yaitu:

Menurut Darmansah, Ni wayan Wardani (2021), Institut Teknologi Telkom Purwekerto, berjudul “Analisis Persebaran Penularan Virus Corona Di Provinsi Jawa Tengah Menggunakan Metode K-Means Clustering”. Dalam penelitian yang mereka lakukan menunjukkan bagaimana menentukan tingkat persebaran *covid-19* di daerah Jawa Tengah dengan menggunakan metode *k-means clustering* dan dibagi dibagi menjadi tiga *cluster*. *Cluster* pertama dengan kategori penularan rendah adalah C0, lalu *cluster* kedua dengan kategori penularan sedang adalah C1, dan yang ketiga dengan kategori penularan tinggi adalah C2. Serta perangkat lunak yang digunakan dalam pengolahan data adalah *rapidminer studio 7.6* [7].

Penelitian selanjutnya dari Desy Noor Permata Saria, YL. Sukestiyarno (2021), Universitas Negeri Semarang, berjudul “Analisis Cluster Metode K-Means pada Persebaran kasus COVID-19 Berdasarkan Provinsi di Indonesia” Dalam penelitian mereka metode *k-means clustering* digunakan untuk mencari data tingkat sebaran kasus *covid-19* dengan hasil *cluster* tinggi, sedang, dan rendah pada setiap provinsi di Indonesia. Dalam penelitiannya mempertimbangkan beberapa aspek jumlah penduduk, kepadatan penduduk, pasien yang positif *covid-19*, pasien yang sembuh, dan pasien yang meninggal dunia. Lalu perangkat lunak yang digunakan untuk melakukan analisis data adalah *software SPSS staristics*[8].

Penelitian lain yang dilakukan oleh Dina Tri Utari (2021), Universitas Islam Indonesia, berjudul “Analisis Karakteristik Wilayah Transmisi Covid-19 Dengan Menggunakan Metode

Clustering” dalam penelitiannya bertujuan untuk mencari informasi kasus positif covid-19 berdasarkan jenis transmisi covid-19 di wilayah bali. Hasil dari penelitian menunjukkan transmisi dapat dikelompokkan menjadi empat *cluster* yang masing-masing memiliki karakteristik, untuk pembatasan transmisi lokal atau perjalanan di dalam negeri termasuk pada *cluster* satu, dua, dan tiga. Sedangkan pembatasan transmisi lokal untuk perjalanan ke luar negeri termasuk pada *cluster* empat[9].

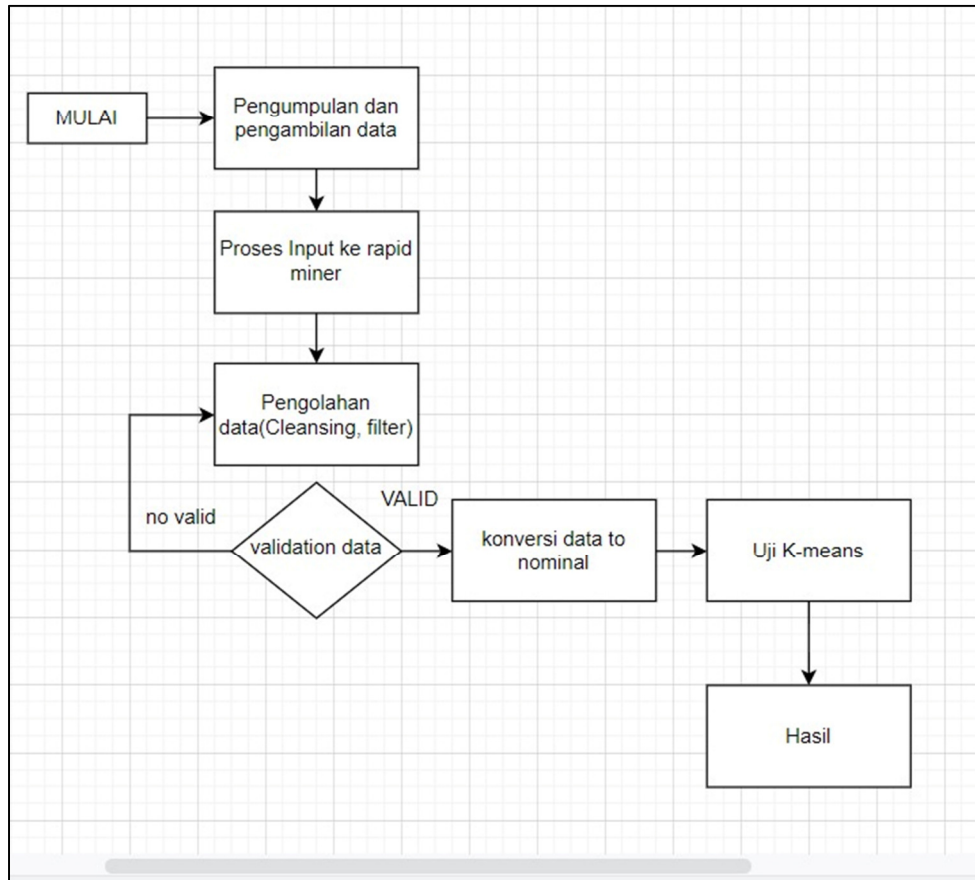
Menurut penelitian Hasrif SY, Rismayani, Asrul Syam(2019), STMIK Dipanegara Makasar, yang berjudul “Data mining menggunakan algoritma K-means pengelompokan penyebaran diare di kota makasar” K-Means Clustering merupakan metode clustering non hirarki yang membagi data kedalam kelompok atau cluster berdasarkan data yang memiliki karakteristik yang sama[10].

Pada penelitian (7–10) telah membahas persebaran covid-19 di indonesia namun ada beberapa kekurangan yang harus diperbaiki yaitu belum menerapkan keputusan zonasi terkait hasil dari pengolahan data. Serta cakupan wilayah yang kurang luas. Maka dari itu jurnal ini dibuat untuk memudahkan pengambilan keputusan zonasi covid-19 dengan menggunakan algoritma K-Means Clustering.

2. METODE PENELITIAN

Metode yang digunakan dalam proses pengolahan data dan referensi dari teknik analisis yang digunakan adalah *clustering* dan dilakukan proses pemilihan data dalam jumlah besar yang bersumber dari *dataset online Kaggle.com*, mengenai penyebaran *covid-19* di Indonesia pada tahun 2020 sampai dengan tahun 2021. Kemudian dilakukan sortasi dan pembersihan agar tersusun sesuai dengan kebutuhan penelitian yaitu dengan 20.683 data dan 2 atribut penelitian.

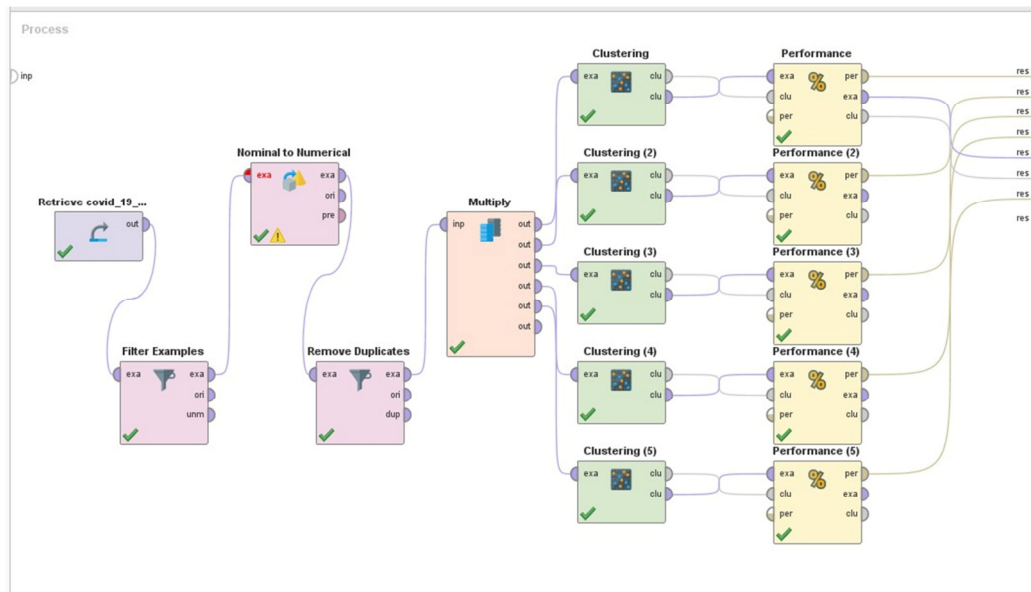
Untuk *software* yang digunakan pada penelitian ini adalah *rapidminer* untuk membantu proses data *mining mulai dari* tahap proses hingga nanti menjadi hasil sehingga dapat membuat keputusan yang paling baik(9).



Gambar 1. Flowchart Penelitian Cluster K-Means

Pada gambar 1 merupakan langkah-langkah atau flowchart dalam penelitian ini, tentu saja langkah awal yang dilakukan adalah mengumpulkan data-data yang nantinya akan diolah serta pengolahan data menggunakan perangkat lunak *rapidminer*. Setelah data valid maka data tersebut akan di uji dengan metode *k-means* yang hasil akhirnya akan berupa *cluster* untuk mengetahui hasil wilayah dalam penelitian ini.

3. HASIL DAN PEMBAHASAN

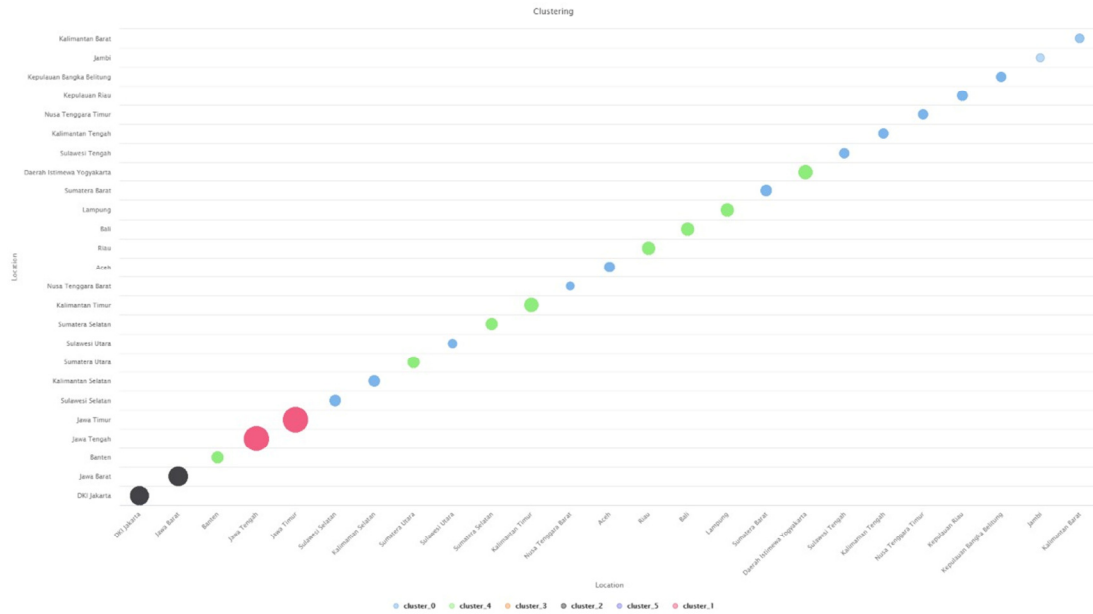


(Sumber: Penulis)

Gambar 1. Proses Metode *Clustering*

Pada gambar 1 merupakan proses yang dilakukan ketika menggunakan metode *clustering* dengan aplikasi *rapid miner* berikut langkah – langkah yang dilakukan :

1. Proses pertama, *dataset* yang masih berupa format *excel* dimasukan ke dalam aplikasi *rapid miner* untuk nantinya dapat dibaca oleh aplikasi.
2. Pada tahap kedua, dilakukan *filter example* yang bertujuan untuk menyaring data meninggal dunia serta membuang data yang tidak digunakan atau tidak *valid*, sehingga nantinya dapat memudahkan dalam pengolahan data.
3. Selanjutnya tahap ketiga, *nominal to numerical* adalah untuk mengkonversi data yang masih format *nominal* diubah ke *numerical*.
4. Lalu pada tahap 4, dilakukan *remove duplicates* untuk menghapus data yang sama atau duplikat, sehingga ketika diproses tidak ada data yang sama atau ganda.
5. Kemudian tahap 5, *multiply* bertujuan untuk membagi pengujian ke setiap *clustering* dengan jumlah *K* yang berbeda, maka dari itu sudah dilakukan pengelompokan sesuai *cluster* masing – masing.
6. Dan yang terakhir, *performance* digunakan untuk mengetahui nilai dari *davies bouldin* yang paling rendah, maka *cluster* tersebut yang akan digunakan sebagai acuan data yang digunakan pada penelitian ini.



(Sumber: Penulis)

Gambar 2. Hasil dari Proses *Clustering*

Terlihat pada gambar 2 hasil dari penelitian yang telah dilakukan terbagi menjadi beberapa *cluster* yaitu:

1. Peringkat pertama, berada di *cluster_01* yaitu daerah Jawa Tengah dan Jawa Timur.
2. Peringkat kedua, berada di *cluster_02* yaitu daerah DKI Jakarta dan Jawa Barat.
3. Peringkat ketiga, berada di *cluster_04* yaitu daerah Banten, Sumatera Utara, Sumatera Selatan, Kalimantan Timur, Riau, Bali, Lampung dan Daerah Istimewa Yogyakarta.
4. Peringkat keempat, berada di *cluster_05* yaitu daerah Sulawesi Selatan, Kalimantan Selatan, Sulawesi Utara, Aceh, Riau, Sumatera Barat, Sulawesi Tengah, Kalimantan tengah, Nusa Tenggara Timur, Kepulauan Riau, Kepulauan Bangka Belitung dan Kalimantan Barat.
5. Peringkat kelima, berada di *cluster_0* yaitu daerah Jambi.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dimulai dari pendahuluan sampai dengan hasil serta pembahasan dan kesimpulan dari penelitian yaitu Alogaritma *K-Means Clustering* dapat diterapkan dalam pembentukan *cluster* wilayah tingkat meninggal dunia akibat *covid-19* khususnya di Indonesia, disimpulkan bahwa hasil dari penelitian menunjukkan *cluster_01* menjadi wilayah dengan tingkat meninggal dunia paling tinggi yaitu daerah Jawa Timur dan Jawa Tengah ditetapkan sebagai zona merah. Maka dari itu pemerintah Indonesia harus meningkatkan kewaspadaan dan prokes pada daerah tersebut.

5. SARAN

Peneliti menyadari keterbatasan pengetahuan yang dimiliki, namun peneliti mencoba memberikan saran yang mungkin nantinya dapat membangun, antara lain:

1. Penelitian ini masih menggunakan satu perangkat lunak yaitu *rapidminer* apabila diolah dengan perangkat lunak yang lain dapat memberikan hasil yang berbeda.
2. Penelitian ini hanya membahas tingkat meninggal dunia yang disebabkan oleh *covid-19*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada teman-teman dan dosen dari STMIK LIKMI yang telah memberi semangat dan dukungan yang membantu terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Unicef. *Coronavirus* [Internet]. 2021 [dikutip 9 April 2022]. Tersedia pada: <https://www.unicef.org/indonesia/id/coronavirus>
- [2] Mulyadi M. *Partisipasi Masyarakat Dalam Penanganan Penyebaran COVID-19*. Bidang Kesejahteraan Sosial Info Singkat. XII(april 2020).
- [3] Binus. Clustering [Internet]. 2021 [dikutip 17 April 2022]. Tersedia pada: <https://socs.binus.ac.id/2017/03/09/clustering/>
- [4] *Definisi Fungsi Metode dan Penerapan Data Mining [Internet]*. 2022 [dikutip 17 April 2022]. Tersedia pada: <https://idcloudhost.com/apa-itu-data-mining-definisi-fungsi-metode-dan-penerapannya/>
- [5] Susanto Sani, Suryadi Dedi. *Pengantar Data Mining*. Nikodemus WK. Andi Yogyakarta; 2010.
- [6] Radius Sarreta I. *Data Mining: Pengertian, Fungsi, Penerapan, dan Metode Pengambilannya*. Cermati.com [Internet]. 2021; Tersedia pada: <https://www.cermati.com/artikel/data-mining-pengertian-fungsi-penerapan-dan-metode-pengambilannya>
- [7] D, Wardani NW. *Analisis Pesebaran Penularan Virus Corona di Provinsi Jawa Tengah Menggunakan Metode K-Means Clustering*. VIII:105–17.
- [8] D. N. P S, Y S. *Analisis Cluster Dengan Metode K-Means pada Persebaran Kasus COVID-19 Berdasarkan Provinsi di Indonesia*, PRISMA, Prosiding Seminar Nasional Matematika,.
- [9] D. T. U. *Analisis Karakteristik Wilayah Transmisi Covid-19 Dengan Menggunakan Metode KMeans Clustering*.

- [10] SY H, Rismayani, Syam A. *Data Mining Menggunakan Algoritma K-Means Pengelompokan Penyebaran Diare di Kota Makassar*. Prosiding Seminar Ilmiah Sistem Informasi dan Teknologi Informasi. 2019;VII:73–82.