

## Analisis Clustering Kasus Covid – 19 di Indonesia Menggunakan Algoritma K-Means

**Angga Aditya Nugroho<sup>1</sup>, Paduloh Paduloh<sup>2\*</sup>**

<sup>1,2</sup>Program Studi Teknik Industri, Fakultas Teknik, Universitas Bhayangkara Jakarta Raya, Jakarta, Indonesia

e-mail: <sup>1</sup> agga.aditya.nugroho19@mhs.ubharajaya.ac.id, <sup>2\*</sup> paduloh@dsn.ubharajaya.ac.id

### Abstract

*The Covid-19 virus is a case that first occurred in Wuhan City, China, which attacks the respiratory system, resembling influenza or flu which attacks the lungs. This research has the aim of being a class analysis to find out how the spread of the virus occurred in Indonesia. This study uses the K – Means method or the median value in calculations based on age in all Indonesian provinces for the last 3 years. The results of the study obtained the highest distribution in the age group 40-60 years in 2020, so it can be said that the decline in Covid - 19 in 2021 has decreased in analysis carried out using the Rapid Miner software*

**Keywords :** RapidMiner, Clustering, Age

### Abstrak

Virus Covid – 19 merupakan kasus yang terjadi pertama kali di Kota Wuhan China yang menyerang pernafasan dengan menyerupai penyakit influenza atau flue yang menyerang paru – paru . Penelitian ini memiliki tujuan sebagai analisis pengkelasan untuk mengetahui bagaimana persebaran virus yang terjadi di Indonesia. Penelitian ini menggunakan metode K – Means atau nilai tengah dalam perhitungan berdasarkan umur di seluruh provinsi Indonesia selama 3 tahun terakhir. Hasil penelitian memperoleh persebaran yang tertinggi terbanyak berada pada kelompok umur 40 – 60 tahun pada tahun 2020 maka dapat dikatakan penurunan Covid – 19 pada tahun 2021 mengalami penurunan analisis yang dilakukan menggunakan software Rapid Miner

**Kata Kunci:** RapidMiner, Clustering, Umur

### PENDAHULUAN

Pada 31 Desember 2019, Kantor WHO di Tiongkok melaporkan kasus penurunan darah yang tidak diketahui penyebabnya di Wuhan, Provinsi Hubei, Tiongkok. Pada 7 Januari 2020, Tiongkok mengidentifikasi Organisasi Kesehatan Dunia (WHO), menamai virus baru sindrom pernapasan akut parah coronavirus-2 (SARS-CoV-2), dan secara resmi menyatakan penyakit itu pada manusia. Kami menyebutnya novel coronavirus (Zhang et al., 2021). Umumnya dikenal sebagai penyakit coronavirus 2019 atau Covid-19. Ketika virus ini menyerang manusia, ia menyebabkan segala sesuatu mulai dari infeksi pernapasan seperti influenza (MERS) dan penyakit terparah merupakan Pernafasan Sindrom (SARS). Sejauh ini, 222 negara telah mengonfirmasi kasus, dengan 127.877.462 kasus terkonfirmasi dan 2.796.561 kematian(Chen et al., 2020).

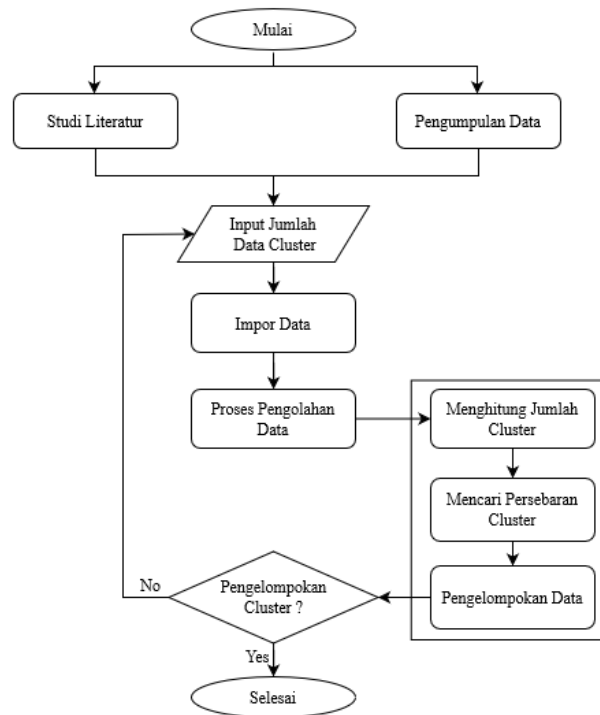
Organisasi Kesehatan Dunia (WHO) memberikan nama *new* virus yakni sindrom pernafasan akut Coronavirus-2 (SARS-CoV-2), dan penyakit ini secara resmi dikenal sebagai penyakit coronavirus 2019 atau Covid-19 pada manusia(Ghosh et al., 2020). Ketika virus ini menyerang manusia, ia menyebabkan segala sesuatu mulai dari infeksi pernapasan seperti influenza hingga penyakit yang lebih serius seperti Sindrom Pernafasan Timur Tengah (MERS) dan Sindrom Pernafasan Akut Parah (SARS). Sejauh ini, 222 negara telah mengonfirmasi kasus, dengan 127.877.462 kasus terkonfirmasi dan 2.796.561 kematian.(Muhammad et al., 2020).

Seiring terjadinya peningkatan kasus covid – 19 di Provinsi Indonesia masyarakat dihimbau untuk menerapkan jaga jarak pada setiap aktivitas yang dilakukan agar dapat menurunkan tingkat penyebarannya(Belostecinic et al., 2022). Karena luasnya daratan Indonesia, kita perlu mengelompokkan bagian-bagian tersebut berdasarkan wilayah Indonesia. Menurut Kementerian Kesehatan Republik Indonesia, jumlah kasus terkonfirmasi positif Covid-19 pada masanya sekitar 1.511.712, jumlah 40.858

kematian dan 1.348.330 sembuh. Indonesia adalah negara dengan jumlah kasus terkonfirmasi tertinggi di ASEAN. Sementara itu, kasus Covid-19 pertama yang dilaporkan di provinsi Indonesia tiba hari itu, dengan jumlah kasus terkonfirmasi positif hingga 10.000. (Stephany et al., 2020).

### METODE PENELITIAN

Metode merupakan seri atau jenis pada penelitian yang sistematiknya berdasarkan angka atau kuantitatif yang pada hakekatnya adalah jenis penelitian yang menggunakan pendekatan induktif dan deduktif (Rosihan et al., 2021). Pendekatan ini diawali dengan kerangka teoritis, gagasan para ahli, dan pemahaman empiris peneliti, berupa usulan pemecahan masalah beserta justifikasinya (verifikasi), atau dukungan data empiris di lapangan evaluasi (Daniel & Harland, 2017) Dari kerangka berpikir diatas menjelaskan mengenai aliran dalam proses penelitian yang dilakukan dengan proses diawali dengan mulai, dilanjutkan dengan studi literatur yang merupakan proses pencarian data melalui beberapa sumber, setelah memperoleh data yang akan di proses, maka akan dilakukan *input* ke dalam *software RapidMiner* dengan rumus *K – Means*, setelah data masuk ke program dilakukan penhelompokan data berdasarkan kelas pada batas umur kasus covid – 19, setelah diketahui pengelompokan data maka diakhiri dengan penyelesaian proses dalam penelitian.



Gambar 1. Kerangka berpikir

### HASIL DAN PEMBAHASAN

Setelah dilakukannya perolehan data dari studi literatur maka menghasilkan data covid – 19 yakni berdasarkan kelompok umur 40 – 60 tahun pada kota yang berada di provinsi indonesia sebanyak 33 provinsi pada tahun 2019 sampai dengan 2021.

Tabel 1 Data kasus Covid - 19

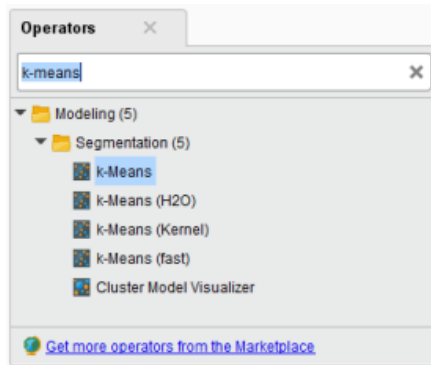
PROVINSI	2019	2020	2021
ACEH	21004	99084	99067
SUMATERA UTARA	17047	99044	99032
SUMATERA BARAT	19071	99061	99061
RIAU	25001	99053	99065
JAMBI	22084	99082	99066

SUMATERA SELATAN	22041	99071	99065
BENGKULU	23098	99078	99076
LAMPUNG	24034	99074	99053
KEP0 BANGKA BELITUNG	29089	99070	99047
KEP0 RIAU	23011	99055	99062
DKI JAKARTA	22055	99064	99053
JAWA BARAT	27039	99066	99050
JAWA TENGAH	34044	99073	99066
DI YOGYAKARTA	31027	99089	99070
JAWA TIMUR	30097	99054	99040
BANTEN	23064	99040	99045
BALI	38058	99057	99070
NUSA TENGGARA BARAT	31023	99052	99052
NUSA TENGGARA TIMUR	31048	98057	98042
KALIMANTAN BARAT	22091	98060	98064
KALIMANTAN TENGAH	31036	99049	99055
KALIMANTAN SELATAN	31052	99048	99038
KALIMANTAN TIMUR	23023	99073	99060
KALIMANTAN UTARA	25056	98094	98067
SULAWESI UTARA	77029	99059	99037
SULAWESI TENGAH	71097	98038	98050
SULAWESI SELATAN	70029	99025	99029
SULAWESI TENGGARA	79039	99010	99016
GORONTALO	23064	98092	99002
SULAWESI BARAT	38058	98033	98019
MALUKU	31023	99050	99052
MALUKU UTARA	31048	99004	98084
PAPUA BARAT	21043	97089	98008
PAPUA	27086	82099	83043

Proses perhitungan data menggunakan *software rapidminer* dengan k – means dan clustering pada data diatas yang bertujuan untuk mengetahui tingkat persebaran serta pengelompokan pada kelompok umur pada provinsi Indonesia sebanyak 33 provinsi dengan data yang digunakan berbentuk format \*.xls.

### Implementasi Menggunakan K – Means

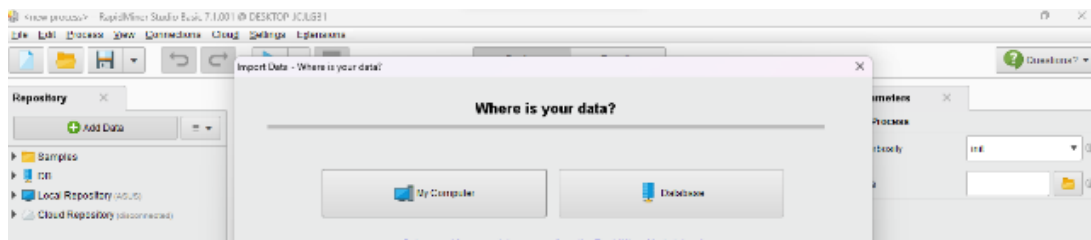
Dalam penerapan K – Means memiliki fungsi agar dapat mengetahui data kasus yang tersebar Covid -19 di Provinsi Indonesia. Dalam tahapan ini merupakan akan dilakukannya pemrosesan data yang telah diperoleh ke dalam *software* RapidMiner yang digunakan sebagai alat untuk mencapai hasil. Pada program dapat dilihat beberapa pilihan tetapi akan dipilih salah satu saja yakni K – Means yang merupakan metode dalam pengujian yang dilakukan saat ini, langkah yang dilakukan dapat dilihat pada penjelasan dibawah.



Gambar 2. Pilihan K – means

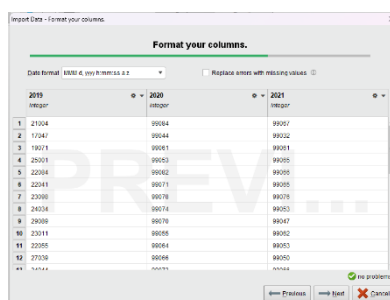
### Import Data

Dalam melakukan pemrosesan diperlukan data berbentuk f\*.xls yang akan dimasukkan pada program dengan cara klik “Add Data”. Selanjutnya memilih lokasi data disimpan yang berbentuk \*.xls pada menu bar “My Computer”.



Gambar 3. Import Data

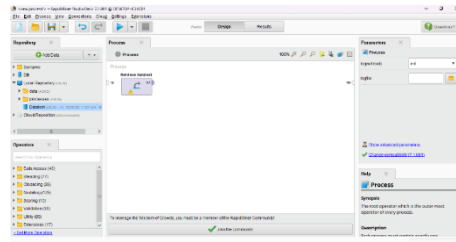
Setelah melakukan pemilihan data serta tidak mengalami permasalahan maka akan terlihat notifikasi seperti “no problems” pada proses import data hingga selesai.



Gambar 4. Proses Import Data

### Membuat Lembar Proses

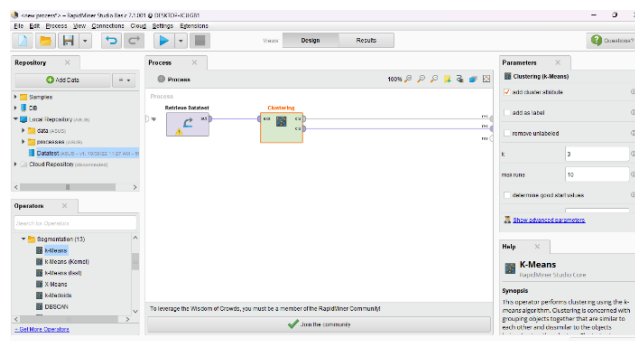
Dalam tahap ini merupakan tahapan dimana akan melakukan proses pencarian persebaran dengan membuat lembaran proses pada program RapidMiner sebelum dilakukannya pengidentifikasian atau untuk mengetahui *clustering* dari data sesuai dengan format yang telah ditentukan agar tidak mengalami masalah pada saat proses dimulai pada tahapan selanjutnya dengan cara di klik dan menyeret ke dalam lembaran proses yang dapat dilihat pada gambar 5.



Gambar 5. Lembar Proses

**Menerapkan Algoritma K – Means**

Pada tahapan ini merupakan dilakukannya data yang telah masuk ke dalam lembar proses dan akan diproses untuk mengetahui kelompok data mana saja yang termasuk dalam kelas persebaran. Yaitu dengan cara mencari “k-Means” pada bagian Operator yang dilanjutkan dengan menarik pada “data” setelah itu kaitkan ke “Clustering” dan “res”. Gambar 6 dibawah.



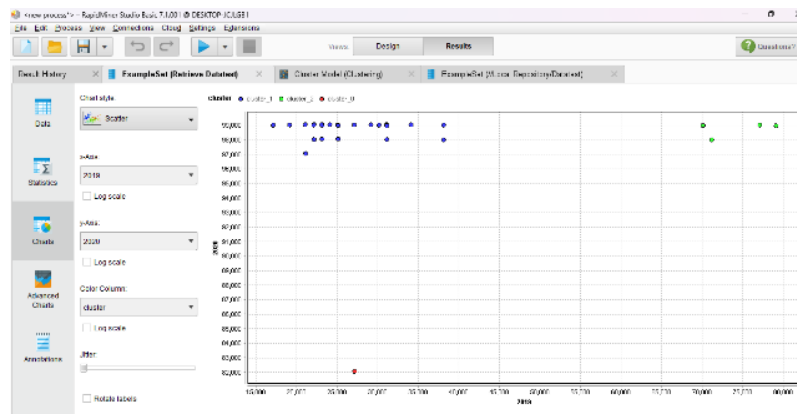
Gambar 6. Menerapkan K - Means

Pada prosesnya akan diinput nilai K sesuai dengan keinginan peneliti yaitu dengan memasukkan 4 nilai, karena dalam penelitian akan melakukan 3 cluster berdasarkan prses algoritma K – Means pada program RapidMiner. Dengan hasil dapat dilihat pada gambar berikut.

RowNo.	Id	cluster	2019	2020	2021
1	1	cluster_1	21004	50034	50017
2	2	cluster_1	17047	50044	50032
3	3	cluster_1	19011	50011	50021
4	4	cluster_1	25001	50051	50035
5	5	cluster_1	22004	50012	50016
6	6	cluster_1	22041	50071	50015
7	7	cluster_1	23032	50072	50076
8	8	cluster_1	24034	50074	50013
9	9	cluster_1	25000	50070	50047
10	10	cluster_1	22011	50010	50012
11	11	cluster_1	22005	50014	50013
12	12	cluster_1	22030	50016	50010
13	13	cluster_1	14044	50073	50016
14	14	cluster_1	31027	50010	50070
15	15	cluster_1	30037	50014	50040
16	16	cluster_1	23004	50040	50045

Gambar 7. Hasil Clustering

Tahapan ini merupakan proses akhir setelah melakukan proses run atau menjalankan proses, jika terdapat tampilan seperti gambar diatas maka dikatakan bahwa proses berjalan dengan baik sehingga dapat ditentukan pengelompokan data yang diproses pada provinsi dan tahun pada data. Setelah diketahui kelompok data maka dapat dilihat pada menu custom untuk mengetahui tampilan visualisasi clustering covid – 19 pada sisi kiri dapat dilihat sumbu x dan y yang menggambarkan persebaran pada level 0 sampai dengan level 3.



Gambar 8. Visualisasi Covid – 19 2020

Dari hasil proses yang dilakukan dalam pengelompokan data atau *clustering* kasus Covid – 19 berdasarkan kelompok umur di Provinsi Indonesia, diketahui bahwa persebaran yang diberikan kelompok pada level 0 sampai 3 yang merupakan batasan pengelompokan. Terjadi persebaran tertinggi dari kelompok umur 40 – 60 pada tahun 2020 mengalami peningkatan yang signifikan sebesar 00,00 - 50,00 dengan warna biru 19 titik peluang kematian dan mengalami penurunan pada tahun 2021 dengan batas 00,00 – 16,00 sebanyak 5 titik peluang kesembuhan dalam program.

## KESIMPULAN DAN SARAN

Berdasarkan pengujian yang dilakukan menggunakan software rapidminer diketahui bahwa persebaran atau clustering yang terbanyak di kelompok umur 40 - 60 tahun berada di posisi cluster 1 pada tahun 2020. Maka dapat dikatakan tingkat covid – 19 terjadi pada tahun 2020 dan mengalami penurunan di tahun 2021. Diketahui penggunaan metode pengelompokan atas persebaran yang dilakukan pada penelitian yakni adalah pemberian kelas kelompok, sehingga dapat mengetahui permasalahan dalam persebarannya.

## DAFTAR PUSTAKA

- Belostecinic, G., Mogoş, R. I., Popescu, M. L., Burlacu, S., Rădulescu, C. V., Bodislav, D. A., Bran, F., & Oancea-Negescu, M. D. (2022). Teleworking—an economic and social impact during covid-19 pandemic: A data mining analysis. *International Journal of Environmental Research and Public Health*, 19(1). <https://doi.org/10.3390/ijerph19010298>
- Chen, T., Peng, L., Yin, X., Jing, B., Yang, J., Cong, G., & Li, G. (2020). A policy category analysis model for tourism promotion in china during the covid-19 pandemic based on data mining and binary regression. *Risk Management and Healthcare Policy*, 13, 3211–3233. <https://doi.org/10.2147/RMHP.S284564>
- Daniel, B. K., & Harland, T. (2017). Higher Education Research Methodology. *Higher Education Research Methodology*. <https://doi.org/10.4324/9781315149783>
- Ghosh, K., Abdul, S., Gayen, S., & Jha, T. (2020). Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information. January.
- Muhammad, L. J., Islam, M. M., Usman, S. S., & Ayon, S. I. (2020). Predictive Data Mining Models for Novel Coronavirus (COVID-19) Infected Patients' Recovery. *SN Computer Science*, 1(4), 1–7. <https://doi.org/10.1007/s42979-020-00216-w>
- Rosihan, R. I., Paduloh, P., & Sulaeman, D. (2021). Penerapan Collaborative Planning, Forecasting and Replenishment (Cpfr) Guna Mengurangi Bullwhip Effect Di Pt.Xyz. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST)*, 2021.
- Stephany, F., Stoehr, N., Darius, P., Neuhäuser, L., Teutloff, O., & Braesemann, F. (2020). The CoRisk-Index: A Data-Mining Approach to Identify Industry-Specific Risk Assessments Related to COVID-19 in Real-Time. *SSRN Electronic Journal*, March, 1–18. <https://doi.org/10.2139/ssrn.3607228>

Zhang, C., Jiang, J., Jin, H., & Chen, T. (2021). The impact of covid-19 on consumers' psychological behavior based on data mining for online user comments in the catering industry in China. *International Journal of Environmental Research and Public Health*, 18(8). <https://doi.org/10.3390/ijerph18084178>

*Angga Aditya Nugroho, Paduloh Paduloh*

Submitted: **19/01/2023**; Revised: **06/06/2023**; Accepted: **27/06/2023**; Published: **30/06/2023**