ANALISIS CLUSTER DENGAN METODE K-MEANS PADA PERSEBARAN KASUS COVID-19 BERDASARKAN PROVINSI DI INDONESIA

Tugas Akhir

Disusun sebagai salah syarat untuk memperoleh gelar Ahli Madya Program Studi Statistika Terapan dan Komputasi

oleh

Desy Noor Permata Sari

4112317029

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI SEMARANG

2020

PERNYATAAN

Dengan ini, saya

Nama : Desy Noor Permata Sari

NIM : 4112317029

Program studi : Statistika Terapan dan Komputasi

Menyatakan bahwa tugas akhir berjudul Analisis Cluster Dengan Metode K-Means Pada Persebaran Kasus Covid-19 Berdasarkan Provinsi Di Indonesia ini bebas dari plagiat. Atas pernyataan ini apabila terbukti plagiat dikemudian hari dalam tugas akhir ini, maka saya bersedia menerima sanksi berdasarkan ketentuan peraturan perundang-undangan.

Semarang, 8 Oktober 2020

Desy Noor Permata Sari

4112317029

Dipindai dengan CamScanner

PENGESAHAN

Tugas Akhir berjudul *Analisis Cluster Dengan Metode K-Means Pada Persebaran Kasus Covid-19 Berdasarkan Provinsi Di Indonesia* karya Desy Noor Permata Sari NIM 4112317029 ini telah dipertahankan dalam Ujian Tugas Akhir FMIPA Universitas Negeri Semarang pada tanggal 8 Oktober 2020 dan disahkan oleh Panitia Ujian.

Semarang, 1 November 2020

Panitia

Ketua,

Dr. Sugianto, M.Si.
NIP 196102191993031001

Sekertaris,

Dr. Mulyono, M.Si.
NIP 197009021997021001

Penguji I,

Drs. Sugiman, B.Sc., M.Si.
NIP 196401111989011001

Penguji II/Pembimbing,

Prof. YL. Sukestiyarno, M.S., Ph.D.
NIP 195904201984031002

MOTTO DAN PERSEMBAHAN

MOTTO

- Bukanlah ilmu yang semestinya mendatangimu, tetapi kamulah yang seharusnya mendatangi ilmu itu (Imam Malik)
- Karunia Allah yang paling lengkap adalah kehidupan yang didasarkan pada ilmu pengetahuan (Ali bin Abi Thalib)
- Ilmu pengetahuan itu bukanlah yang dihafal, melainkan yang memberi manfaat (Imam Syafi'i)

PERSEMBAHAN

- Untuk Allah SWT.
- Untuk Ayah, Ibu, dan Kakak.
- Untuk sahabat-sahabat.
- Untuk teman-teman Adio Dance Crew Semarang.
- Untuk teman-teman Staterkom Angkatan 2017.
- Untuk teman-teman Organisasi Mipa English Club.
- Untuk seluruh dosen Jurusan Matematika.

PRAKATA

Segala Puji syukur atas kehadirat Allah SWT yang telah melimpahkan rahmat dan hidayah- Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik dan lancar. Tugas Akhir ini berjudul "**Analisis Cluster Dengan Metode K-Means Pada Persebaran Kasus Covid-19 Berdasarkan Provinsi Di Indonesia**" disusun sebagai salah satu syarat untuk memperoleh gelar Diploma (D3) pada Program Studi Statistika Terapan dan Komputasi Fakultas Matematika dan Ilmu Pengetahuan Alam.

Dalam mengerjakan dan menyusun Tugas Akhir ini, penulis telah banyak mendapat bantuan, bimbingan, dorongan, dan petunjuk yang sangat bermanfaat dari berbagai pihak yang sangat mendukung. Oleh karena itu pada kesempatan ini penulis mengucapkan terima kasih dengan tulus kepada:

1. Prof. Dr. Fathur Rokhman, M.Hum., Rektor Universitas Negeri Semarang.
2. Dr. Sugianto, M.Si., Dekan FMIPA UNNES.
3. Dr. Mulyono, M.Si., Ketua Jurusan Matematika FMIPA UNNES.
4. Dr. Iqbal Kharisudin, M.Sc., Koordinator Prodi Statistika Terapan dan Komputasi FMIPA UNNES.
5. Prof. YL. Sukestiyarno, M.S., Ph.D., Dosen pembimbing yang telah memberikan bimbingan, pengarahan, dan motivasi kepada penulis dalam penyusunan Tugas Akhir ini.
6. Drs. Sugiman, B.Sc., M.Si., Dosen penguji yang telah membantu dalam kelancaran ujian dan memberikan masukkan dalam penulisan Tugas Akhir.
7. Bapak dan Ibu Dosen Jurusan Matematika yang telah memberikan bekal ilmu kepada penulis dalam penyusunan Tugas Akhir ini.
8. Kedua orang tua dan keluarga besar yang telah membantu, memberi semangat, serta doa yang tanpa henti.
9. Seluruh teman-teman Staterkom Angkatan 2017.
10. Seluruh teman-teman mahasiswa jurusan Matematika.
11. Mbak Putri, Mbak Tita, Thasya, dan Ulfa yang selalu memberikan dukungan.

12. Arum, Aulia, Sasnia, dan Priska yang selalu memberikan semangat.
13. Teman-teman dari organisasi Mipa English Club atas suka dan dukanya.
14. Keluarga Adio Dance Crew yang selalu membantu setiap saat.
15. Semua pihak yang telah membantu baik langsung maupun tidak langsung sehingga Tugas Akhir ini dapat terselesaikan.

Semoga seluruh bantuan, doa, dan dukungan yang telah diberikan kepada penulis dapat bermanfaat untuk ke depannya dan mendapatkan balasan yang terbaik dari Allah SWT.

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini terdapat banyak kekurangan. Maka dari itu, penulis mengharapkan saran dan kritik yang membangun. Akhir kata, penulis mengharapkan semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi penulis dan pembaca.

Semarang, 8 Oktober 2020

Desy Noor Permata Sari

ABSTRAK

Sari, Desy Noor Permata. (2020). Analisis Cluster Dengan Metode K-Means Pada Persebaran Kasus Covid-19 Berdasarkan Provinsi Di Indonesia. Tugas Akhir. Prodi Statistika Terapan dan Komputasi, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang, Pembimbing: Prof. YL. Sukestiyarno, M.S.,Ph.D.

Kata Kunci: Artificial Intelligence, K-Means, COVID-19, Analisis Cluster

Industri 4.0 merupakan perubahan besar dari inovasi di bidang Elektronik dan Teknologi Informasi, salah satunya *Artificial Intelligence*. *Artificial Intelligence* (AI) dapat belajar mendeteksi wabah dan memperkirakan resiko puncak dari penyakit menular. *K-Means Cluster* adalah clustering non-hierarki yang mengelompokkan data ke dalam suatu cluster yang memiliki karakteristik sama ke dalam satu cluster yang sama. Hal yang dapat dilakukan untuk menangani penyebaran COVID-19 yang makin meluas yaitu dengan menganalisis laporan berita, *sentimen* masyarakat di media sosial, dan juga dokumen yang dirilis pemerintah tentang persebaran COVID-19.

Tujuan dalam penelitian ini menggunakan metode *K-Means Cluster* untuk mengetahui tingkat persebaran kasus COVID-19 kategori tinggi, sedang, dan rendah pada masing-masing Provinsi di Indonesia. Ada beberapa aspek yang bisa diukur seperti jumlah penduduk, kepadatan penduduk, kasus positif terinfeksi COVID-19, pasien yang sembuh, dan pasien yang meninggal dunia.

Metode pengumpulan data yang digunakan adalah metode dokumentasi berupa data sekunder yang diperoleh dari publikasi buku Badan Pusat Statistik dan data Kemenkes RI di Badan Nasional Penanggulangan Bencana. Data yang digunakan yaitu jumlah penduduk (X_1), kepadatan penduduk (X_2), positif (X_3), sembuh (X_4), dan meninggal (X_5) dan dianalisis menggunakan software SPSS.

Dari hasil penelitian dengan metode K-Means terbentuk menjadi 5 cluster. Cluster 1 termasuk kasus yang tinggi berisi 2 Provinsi. Cluster 2 termasuk kasus yang sedang berisi 3 Provinsi. Cluster 3 termasuk kasus yang rendah berisi 29 Provinsi dan dibagi lagi menjadi 3 cluster dengan mengelompokkan berdasarkan tingakatannya. Karakteristik cluster 1 kategori tinggi berisi rata-rata X_2, X_3, X_4, dan X_5 berada di atas rata-rata populasi. Cluster 2 kategori sedang berisi rata-rata X_2, X_3, X_4, dan X_5 berada di bawah rata-rata populasi. Cluster 3 kategori rendah tingkat pertama berisi rata-rata semua variabel berada di atas rata-rata populasi. Cluster 4 kategori rendah tingkat kedua berisi rata-rata X_1 dan X_3 berada di atas rata-rata populasi. Cluster 5 kategori rendah tingkat ketiga berisi rata-rata X_3, X_4, dan X_5 berada di bawah rata-rata populasi. Variabel yang memberikan perbedaan paling besar adalah variabel kepadatan penduduk dengan nilai F sebesar 26,641 dan nilai signifikan 0,000. Provinsi yang memiliki nilai paling besar pada variabel kepadatan penduduk adalah Provinsi DKI Jakarta.

DAFTAR ISI

ANALISIS CLUSTER DENGAN METODE K-MEANS PADA PERSEBARAN KASUS COVID-19 BERDASARKAN PROVINSI DI INDONESIA 1
PERNYATAAN i
PENGESAHAN ii
MOTTO DAN PERSEMBAHAN iii
PRAKATA iv
ABSTRAK vi
DAFTAR ISI vii
DAFTAR TABEL x
DAFTAR GAMBAR xi
LAMPIRAN xii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Batasan Masalah 4
1.4 Tujuan Penelitian 4
1.5 Manfaat Penelitian 5
1.6 Sistematika Penulisan Tugas Akhir 5
BAB II TINJAUAN PUSTAKA 7
2.1 Jumlah Penduduk 7
2.2 Kepadatan Penduduk 7
2.3 Persebaran Kasus COVID-19 7
2.3.1 Positif COVID-19 7
2.3.2 Pasien Yang Sembuh 8
2.3.3 Pasien Yang Meninggal Dunia 8
2.4 Analisis Statistika 8

2.5 Statistika Multivariate 9
2.5.1 Jenis Data Dalam Analisis Multivariat 9
2.6 K-Means Clustering 10
2.7 Kerangka Berpikir 12
BAB III METODE PENELITIAN 15
3.1 Populasi Penelitian 15
3.2 Variabel Penelitian 15
3.3 Metode Pengumpulan Data 15
3.4 Analisis Data Dan Alur Penelitian 16
3.4.1 Analisis Data 16
3.4.2 Alur Penelitian 20
BAB IV HASIL DAN PEMBAHASAN 21
4.1 Hasil Penelitian 21
4.1.1 Data 21
4.1.2 Uji Multikolinearitas 21
4.1.3 Analisis Deskriptif 24
4.1.4 Analisis Cluster dengan K-Means 24
4.1.5 Melihat Tingkat Signifikansi Antar Cluster 30
4.1.6 Jumlah Anggota Di Masing-Masing Cluster 31
4.1.7 Profiling Komposisi Cluster 32
4.1.8 Pengelompokkan Provinsi Pada Cluster 3 37
4.1.9 Diagram Hasil Analisis K-Means Cluster 38
4.2 Pembahasan 39
BAB V PENUTUP 42
5.1 Simpulan 42
5.2 Saran 44

6 DAFTAR PUSTAKA 45
7 LAMPIRAN 47

DAFTAR TABEL

Tabel 4.1 Hasil Uji Multikolinieritas 22
Tabel 4.2 Hasil Uji Multikolinieritas Dengan Transformasi Data 23
Tabel 4.3 Statistik Deskriptif 24
Tabel 4.4 Initial Cluster Centers 25
Tabel 4.5 Iteration History 25
Tabel 4.6 Final Cluster Centers 26
Tabel 4.7 ANOVA 30
Tabel 4.8 Number of Cases in Each Cluster 32
Tabel 4.9 Provinsi Pada Setiap Cluster 34
Tabel 4.10 Statistik Deskriptif Pada Cluster 3 35
Tabel 4.11 Provinsi Pada Setiap Cluster 37

DAFTAR GAMBAR

Gambar 2.1 Kerangka Berpikir 14
Gambar 3.1 Linear Regression 16
Gambar 3.2 Linear Regression Statistics 17
Gambar 3.3 Compute Variable 17
Gambar 3.4 Descriptives 18
Gambar 3.5 K-Means Cluster Analysis 18
Gambar 3.6 K-Means Cluster: Save 19
Gambar 3.7 K-Means Cluster: Options 19
Gambar 3.8 Langkah-langkah Analisis K-Means Cluster 20
Gambar 4.1 Rata-Rata Cluster Pada Masing-Masing Variabel**Error! Bookmark not defined.**
Gambar 4.2 Provinsi Yang Memiliki Kasus Tertinggi Pada Setiap Cluster 35
Gambar 4.3 Hasil Tafsiran Angka Untuk Masing-Masing Variabel Pada Cluster 3 36
Gambar 4.4 Diagram Hasil Analisis K-Means Cluster 38

LAMPIRAN

Lampiran 1 **Data Tingkat Kasus Persebaran COVID-19 Tanggal 31 Mei 2020 Provinsi Di Indonesia.**..... 47

Lampiran 2 **Hasil Ouput SPSS Sebelum Diuji Dengan Analisis *K-Means Cluster*** 49

Lampiran 3 **Hasil Ouput SPSS Analisis *K-Means Cluster*** 50

Lampiran 4 **Hasil Ouput SPSS Analisis *K-Means Cluster* Pada Cluster 3**..... 54

BAB I
PENDAHULUAN

1.1 Latar Belakang

Peran Data Science di *Industri 4.0* untuk tahun 2020 sangat dibutuhkan dan memberikan dampak yang sangat kuat dalam kehidupan ekosistem dunia. *Industri 4.0* merupakan sebuah revolusi dari inovasi di bidang Elektronik dan Teknologi Informasi. Konsep *Industri 4.0* adalah realitas baru dari ekonomi modern, karena inovasi dan pengembangan teknologi memainkan peran penting dalam setiap organisasi (Ślusarczyk, 2018). Terdapat beberapa macam teknologi yang akan menjadi pilar utama dalam revolusi *Industri 4.0*, yaitu *Big Data, Internet of Things*, *Cloud Computing*, *Addictive Manufacturing*, dan salah satunya *Artificial Intelligence*.

Artificial Intelligence (AI) adalah domain Sains dan Teknik yang berkaitan dengan teori dan praktik pengembangan sistem yang menunjukkan karakteristik yang terkait dengan kecerdasan dalam perilaku manusia, seperti persepsi, pemrosesan bahasa alami, pemecahan masalah dan perencanaan, pembelajaran dan adaptasi, dan bertindak atas lingkungan (Tecuci, 2012). *Artificial Intelligence* dalam ilmu teknologi terbagi dalam 7 cabang, yaitu *Machine Learning (ML)*, *Natural Languange Processing (NLP)*, *Expert System*, *Vision*, *Speech*, *Planning*, dan *Robotic*. *Machine Learning* adalah bagian dalam *Artificial Intelligence* dengan kemampuan mesin untuk mengakses data yang ada dengan perintah mereka sendiri. *Machine Learning* dapat mempelajari data yang ada dan melaksanakan tugas-tugas tertentu.

Machine Learning pertama kali diperkenalkan jauh sebelum adanya komputer oleh beberapa ilmuwan matematika terkenal seperti Thomas Bayes dan Pierre-Simon Laplace pada tahun 1812, Adrien-Marie Legendre pada tahun 1805, dan Andrey Markov pada tahun 1913. Tugas-tugas yang dilaksanakan oleh *Machine Learning* berbagai macam. *Machine learning* mempunyai 3 macam teknik pada umumnya, yaitu *Supervised Learning* yang mempelajari pada data yang ada

dan mencari pola dari data set yang akan dijadikan sebagai acuan untuk data-data selanjutnya, *Reinforcement Learning* yang menempatkan agent ke sebuah environment yang tidak diketahui tempatnya dan akan mengeksplorasi keseluruhan sampai menemukan istilah reward dan error, dan *Unsupervised Learning* bersifat deskriptif dan berguna untuk mengkategorikan atau mengelompokkan data (Rauhan, 2019).

Analisis Cluster merupakan teknik multivariate yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis Cluster mengklasifikasi objek sehingga objek-objek yang paling dekat kesamaanya dengan objek lain berada dalam cluster yang sama (Ediyanto et al., 2013). Ciri-ciri pengelompokkan yang baik adalah mempunyai kesamaan (homogenitas) yang tinggi antar anggota dalam satu cluster (within cluster), dan mempunyai ketaksamaan (heterogenitas) yang tinggi antar cluster (between cluster) (Yulianto & Hidayatullah, 2014).

Sekitar awal tahun 2020, dunia gempar dengan penyebaran virus baru yaitu COVID-19 atau bisa disebut dengan Novel coronavirus. Virus tersebut masih satu keluarga dengan SARS-CoV dan MERS-CoV, yakni dua virus yang menimbulkan gangguan pada saluran pernapasan seperti pneumonia. Kini, status COVID-19 menjadi pendemik di sebagian besar penjuru dunia. Penyebaran awalnya berasal dari Wuhan, China, namun menyebar dengan sangat cepat di beberapa negara lainnya, seperti Korea Selatan yang tidak jauh dari China. Tidak hanya itu, penyebaran virus ini juga semakin luas, bahkan ke negara-negara lain yang terletak jauh, seperti Italia dan Iran. Hingga saat ini pun, vaksin untuk menyembuhkan COVID-19 masih dalam penelitian, dimana para peneliti terus berpacu dengan waktu dikarenakan makin banyaknya orang yang terinfeksi atau bahkan mengalami kematian diakibatkan keganasan virus ini.

Ada beberapa hal yang dapat dilakukan untuk menangani penyebaran COVID-19 yang makin meluas diantaranya dengan menganalisis laporan berita, *sentimen* masyarakat di media sosial, dan juga dokumen yang dirilis pemerintah tentang persebaran COVID-19. Artificial Intelligence (AI) dapat belajar mendeteksi

wabah dan memperkirakan resiko puncak dari penyakit menular. BlueDot merupakan salah satu contoh dan suatu startup yang berasal dari Kanada yang menyediakan layanan berbasis Artificial Intelligence (AI) untuk memberi peringatan ancaman penyebaran wabah beberapa hari sebelum pusat pengendalian pencegahan penyakit atau organisasi kesehatan dunia mengeluarkan peringatan publik mereka.

Dalam penelitian ini akan dianalisis kasus pengelompokkan *K-Means Cluster* 34 Provinsi di Indonesia berdasarkan tingkat persebaran kasus COVID-19. *K-Means Cluster* adalah salah satu metode dan clustering non-hierarki yang berusaha mengelompokkan data ke dalam suatu cluster sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu cluster yang sama. Ukuran kesamaan yang digunakan adalah ukuran jarak antar objek. Kedua objek yang memiliki jarak paling dekat akan bergabung menjadi satu cluster. Kedekatan jarak yang dimiliki menunjukkan bahwa kedua objek tersebut memiliki tingkat kesamaan karakteristik.

Untuk mengetahui tingkat persebaran kasus COVID-19 kategori tinggi, sedang, dan rendah pada masing-masing Provinsi di Indonesia. Ada beberapa aspek yang bisa diukur seperti jumlah penduduk, kepadatan penduduk, kasus positif terinfeksi COVID-19, pasien yang sembuh, dan pasien yang meninggal dunia. Berdasarkan tingkat persebaran kasus COVID-19 akan dilakukan clustering pada Provinsi di Indonesia pada tanggal 31 Mei 2020 dengan metode *K-Means Cluster*.

1.2 Rumusan Masalah

Masalah yang dikaji dalam Tugas Akhir ini adalah sebagai berikut.

1. Bagaimana penerapan metode *K-Means Cluster* untuk mengelompokkan provinsi?
2. Bagaimana karakteristik cluster yang terbentuk berdasarkan tingkat persebaran kasus COVID-19 di Indonesia pada tanggal 31 Mei 2020?

3. Variabel apa yang memberikan perbedaan paling besar pada cluster yang terbentuk dalam persebaran kasus COVID-19 dan Provinsi mana yang memiliki nilai paling besar pada variabel ini?

1.3 Batasan Masalah

Agar mendekati tujuan yang diharapkan, maka diperlukan pembatasan masalah sebagai berikut.

1. Dibatasi pada analisis cluster K-Means beserta teori-teori yang mendukung.
2. Studi kasus penelitian ini dibatasi yaitu data kasus COVID-19 berdasarkan provinsi pada tanggal 31 Mei 2020, jumlah penduduk, dan kepadatan penduduk tahun 2020 di Provinsi Indonesia.
3. Analisis data dilakukan dengan data yang diperoleh dari publikasi buku Badan Pusat Statistik di 34 Provinsi dan laporan data Kemenkes dari Badan Nasional Penanggulangan Bencana.
4. Data yang digunakan merupakan data sekunder berupa variabel jumlah penduduk, kepadatan penduduk, positif, sembuh, dan meninggal.
5. Perangkat lunak sebagai alat bantu dalam penelitian yaitu berupa Software IBM SPSS Statistics.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai melalui penelitian ini adalah sebagai berikut.

1. Menerapkan metode *K-Means Cluster* untuk mengelompokkan provinsi dan mengetahui karakteristik cluster yang terbentuk berdasarkan tingkat persebaran kasus COVID-19 di Indonesia pada tanggal 31 Mei 2020.
2. Mengetahui Provinsi yang termasuk dalam cluster tingkat persebaran kasus COVID-19 tinggi, sedang, dan rendah.
3. Mengetahui variabel yang memberikan perbedaan paling besar pada cluster yang terbentuk dan Provinsi mana yang memiliki jumlah terbanyak dalam variabel tersebut.

1.5 Manfaat Penelitian

1. Bagi Pembaca, dapat dijadikan sebagai bahan referensi supaya mentaati dan menerapkan PSBB untuk lebih waspada dan peduli kepada lingkungan sekitar terutama yang berada dalam zona atau cluster dengan memiliki kasus persebaran COVID-19 yang tinggi, serta tidak menganggap remeh dalam kasus COVID-19 ini.
2. Bagi Bidang Kesehatan, dapat dijadikan sebagai bahan masukan untuk meningkatkan pelayanan kesehatan yang lebih baik pada tingkat persebaran kasus COVID-19 pada masing-masing Provinsi di Indonesia tahun 2020.
3. Bagi Pemerintah, dapat dijadikan sebagai bahan masukan untuk pemerintah dalam meningkatkan pengawasan dan penanganan kepada masyarakat khususnya pada daerah yang tingkat persebaran kasus COVID-19 tinggi

1.6 Sistematika Penulisan Tugas Akhir

Sistematika pada penulisan ini digunakan untuk memudahkan dari segi pemahaman jalan pikiran secara keseluruhan dalam tugas akhir. Secara garis besar tugas akhir di sini dibagi menjadi tiga bagian, yakni:

1. Bagian Awal

   Bagian awal pada tugas akhir ini terdiri dari halaman judul, halaman pernyataan, pengesahan, motto dan persembahan, prakata, abstrak, daftar isi, daftar tabel, daftar gambar, dan daftar lampiran.

2. Bagian Isi

   Bagian isi pada tugas akhir ini terdiri dari lima bab. Adapun lima bab tersebut adalah sebagai berikut.

   BAB I PENDAHULUAN

Pada bab ini terdiri atas latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan tugas akhir.

BAB II TINJAUAN PUSTAKA

Pada bab ini diuraikan tentang konsep-konsep yang berisi teori-teori sebagai kerangka berfikir untuk menyelesaikan masalah dalam penelitian tugas akhir ini.

BAB III METODE PENELITIAN

Pada bab ini terdiri atas metode yang digunakan dalam penelitian seperti populasi penelitian, variabel penelitian, metode pengumpulan data, analisis data dan alur penelitian.

BAB IV HASIL DAN PENELITIAN

Pada bab ini menjelaskan tentang hasil penelitian dan pembahasan dari hasil penelitian.

BAB V PENUTUP

Pada bab ini berisi tentang simpulan dan saran- saran dalam penelitian.

3. Bagian Akhir

Bagian akhir tugas akhir ini berisi tentang daftar pustaka dan lampiran-lampiran.

BAB II
TINJAUAN PUSTAKA

2.1 Jumlah Penduduk

Penduduk adalah semua orang yang berdomisili di wilayah geografis Indonesia selama enam bulan atau lebih dan atau mereka yang berdomisili kurang dari enam bulan tetapi bertujuan menetap. Pertumbuhan Penduduk adalah keseimbangan yang dinamis antara kekuatan-kekuatan yang menambah dan kekuatan-kekuatan yang mengurangi jumlah penduduk, yang secara terus menerus akan dipengaruhi oleh jumlah bayi yang lahir (menambah jumlah penduduk), tetapi secara bersamaan pula akan dikurangi oleh jumlah kematian yang terjadi pada semua golongan umur (Pancasa et al., 2018). Kuantitas atau jumlah penduduk dapat sebagai potensi maupun menjadi beban bagi suatu negara, akan menjadi potensi apabila jumlah penduduk seimbang dengan sumber daya yang lain serta mempunyai kualitas hidup yang baik. Sebaliknya, menjadi beban apabila jumlah penduduk melampaui kapasitas wilayah Negara tersebut (Christiani et al., 2013).

2.2 Kepadatan Penduduk

Kepadatan penduduk adalah perbandingan antara jumlah penduduk dengan luas wilayah yang dihuni (Mantra, 2011). Kepadatan penduduk dapat mempengaruhi kualitas hidup penduduknya. Pada daerah dengan kepadatan yang tinggi, usaha peningkatan kualitas penduduk akan lebih sulit dilakukan (Christiani et al., 2013). Daerah yang memiliki kepadatan penduduk yang tinggi akan menyebabkan transmisi penyakit menular lebih cepat dengan rantai penyebaran yang lebih kompak dan kompleks. Maka dari itu, diperlukan kesadaran yang tinggi oleh masyarakat untuk mematuhi segala peraturan yang ada untuk kelangsungan hidup dan kebersamaan yang lebih baik.

2.3 Persebaran Kasus COVID-19

2.3.1 Positif COVID-19

COVID-19 dapat menyebar melalui percikan air (droplet) dari hidung atau mulut pada saat berbicara, batuk atau bersin. Droplet tersebut kemudian jatuh pada

benda di sekitarnya. Kemudian jika ada orang lain menyentuh benda yang sudah terkontaminasi dengan droplet tersebut, lalu orang itu menyentuh mata, hidung atau mulut (segitiga wajah), maka orang itu dapat terinfeksi virus Corona atau bisa juga seseorang terinfeksi virus Corona ketika tanpa sengaja menghirup droplet dari penderita. Gejala-gejala orang yang terinfeksi atau positif COVID-19, yaitu demam $\geq$ 38℃, batuk kering, sesak napas, nyeri tenggorokan, dan pegal-pegal atau merasa kelelahan (Kemenkes RI, 2020).

2.3.2 Pasien Yang Sembuh

Pasien yang terinfeksi COVID-19 dapat sembuh jika menerima perawatan yang tepat untuk meredakan dan mengobati gejala, dan mereka yang sakit serius (gejala sedang dan berat) harus dibawa ke rumah sakit. Sebagian besar pasien sembuh karena perawatan untuk gejala yang dialami (Kemenkes RI, 2020).

2.3.3 Pasien Yang Meninggal Dunia

Usia lanjut dan penyakit komorbid seperti diabetes dan hipertensi dilaporkan menjadi faktor risiko kematian pada orang-orang yang terjangkit COVID-19. Karena itu, lansia berisiko paling tinggi meninggal dan menjadi salah satu kelompok yang paling rentan. Penting untuk diketahui bahwa lansia memiliki hak yang sama dengan orang lain untuk menerima perawatan kesehatan berkualitas tinggi, termasuk perawatan intensif. Pasien-pasien ini bisa saja menunjukkan gejala-gejala ringan tetapi keadaannya berisiko tinggi menjadi lebih buruk dan perlu dirawat di unit yang ditunjuk agar dapat dimonitor dengan baik (World Health Organization, 2020).

2.4 Analisis Statistika

Berdasarkan jumlah variabel yang dianalisis, analisis statistika dapat dikelompokkan menjadi analisis univariat, bivariat, dan multivariate. Analisis univariat berasal dari kata uni dan variate yang berarti analisis satu variabel. Analisis bivariat berasal dari kata bi dan variate berarti analisis dua variabel. Apabila analisis lebih dari dua variabel dinamakan analisis multivariate. Analisis

multivariate adalah perluasan dari analisis univariat dan bivariate (Santoso, 2014). Pada penelitian ini, analisis statistika yang digunakan merupakan analisis multivariate.

2.5 Statistika Multivariate

Secara umum, analisis multivariate atau metode multivariate berhubungan dengan metode-metode statistik yang secara bersama-sama (simultan) melakukan analisis terhadap lebih dari dua variabel pada setiap obyek atau orang. Jadi bisa dikatakan analisis multivariate merupakan perluasan dari analisis univariat (seperti uji t) atau bivariat (seperti metode korelasi dan regresi sederhana). Analisis multivariate adalah analisis multi variabel dalam satu atau lebih hubungan. Analisis ini berhubungan dengan semua teknik statistik yang secara simultan menganalisis sejumlah pengukuran pada individu atau obyek. Beberapa ahli lain mengatakan bahwa tujuan analisis multivariate adalah mengukur, menerangkan, dan memprediksi tingkat relasi diantara variat-variat. Jadi, karakter multivariate tidak sekedar berada pada jumlah variabel atau observasi yang dilibatkan dalam analisis, tetapi juga pada kombinasi berganda antar variat (Simamora, 2005). Metode statistika multivariate yang digunakan dalam penelitian ini adalah metode interpedensi dengan analisis cluster berupa *K-Means Cluster*ing.

Jenis Data Dalam Analisis Multivariat

Menurut (Sukestiyarno, 2016), berdasarkan jenis perolehannya atau pengumpulannya data diklasifikasikan menjadi 5 macam tipe atau skala. Hal tersebut adalah:

1. Skala Nominal

Data nominal diperoleh dari hasil pengamatan (observasi). Apabila datanya disimbolkan menjadi data numerik (kuantitatif) maka bilangan yang digunakan bersifat diskrit dan tidak mengenal urutan. Artinya tiap unsurnya tidak mempunyai arti menurut besarnya atau posisinya.

2. Skala Ordinal

Data ordinal berasal dari hasil pengamatan, observasi, atau angket berskala dari suatu variabel. Hasil observasi berbentuk data kualitatif. Hasil observasi berbentuk data kualitatif. Apabila datanya disimbolkan menjadi data numerik maka bilangan yang digunakan bersifat diskrit dan mengenal urutan menurut kualitas atributnya.

3. Skala Kardinal

Data kardinal berasal dari hasil membilang atau menghitung suatu variabel. Data berbentuk kuantitatif bilangan diskrit, umumnya dinyatakan dalam bilangan kardinal.

4. Skala Interval

Data interval berasal dari hasil mengukur suatu variabel. Data diasumsikan berbentuk bilangan kontinu mempunyai ukuran urutan, seperti data dengan ordinal. Pada skala interval tidak memiliki nol mutlak, artinya jika suatu responden variabelnya bernilai nol bukan berarti tidak memiliki substansi sama sekali.

5. Skala Ratio

Data rasio berasal dari hasil mengukur suatu variabel. Data diasumsikan berbentuk bilangan kontinu hampir sama dengan skala interval, perbedaannya terletak pada nilai nol. Pada skala rasio memiliki nilai nol mutlak, artinya jika suatu responden variabelnya bernilai nol berarti tidak memiliki substansi sama sekali.

2.6 K-Means Clustering

Cluster Analysis merupakan salah satu metode objek mining yang bersifat tanpa Latihan (unsupervised analysis), sedangkan *K-Means Cluster* Analysis merupakan salah satu metode cluster analysis non hirarki yang berusaha untuk

mempartisi objek yang ada kedalam satu atau lebih cluster atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokan dalam satu cluster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokan ke dalam cluster yang lain. Tujuan pengelompokan adalah untuk meminimalkan objective function yang di set dalam proses clustering, yang pada dasarnya berusaha untuk meminimalkan variasi dalam satu cluster dan memaksimalkan variasi antar cluster (Ediyanto et al., 2013). Metode K-Means berusaha mengelompokkan data yang ada kedalam satu kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain (Helilintar & Farida, 2018). Dasar algoritma k-means adalah sebagai berikut:

1. Tentukan nilai k sebagi jumlah cluster yang ingin dibentuk.
2. Inisialisasi k sebagai centroid yang dapat dibangkitkan secara random.

   $d(i,k) = \sqrt{\sum_i^m (C_i j - C_k j)^2}$

3. Hitung jarak setiap data ke masing-masing centroid menggunakan persamaan Eulidiean Distance yaitu sebagai berikut:

   $min \sum_k^i - a_{ik} - = \sqrt{\sum_i^m (C_i j - C_k j)^2}$

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centroidnya.

   $C_{kj} = \frac{\sum_k^i x_{ij}}{p}$

5. Tentukan posisi centroid baru (k)
6. Kembali ke langkah 3 jika posisi centroid baru dengan centroid lama tidak sama.

Menurut (Santoso, 2014), *K-Means Cluster*ing memiliki asumsi-asumsi yang harus dipenuhi, yaitu:

1. Sampel yang diambil harus benar-benar bisa mewakili populasi yang ada.
2. Multikolinearitas, yakni kemungkinan adanya korelasi antar objek. Namun sebaiknya tidak ada, bila ada besar multikolinearitas tersebut tidaklah tinggi (misal di atas 0,5).

Pada penelitian ini menggunakan metode *K-Means Cluster* untuk mengetahui dan mengelompokkan tingkat persebaran kasus COVID-19 pada masing-masing Provinsi di Indonesia, ada beberapa aspek yang bisa diukur seperti jumlah penduduk, kepadatan penduduk kasus positif terinfeksi COVID-19, pasien yang sembuh, dan pasien yang meninggal dunia.

2.7 Kerangka Berpikir

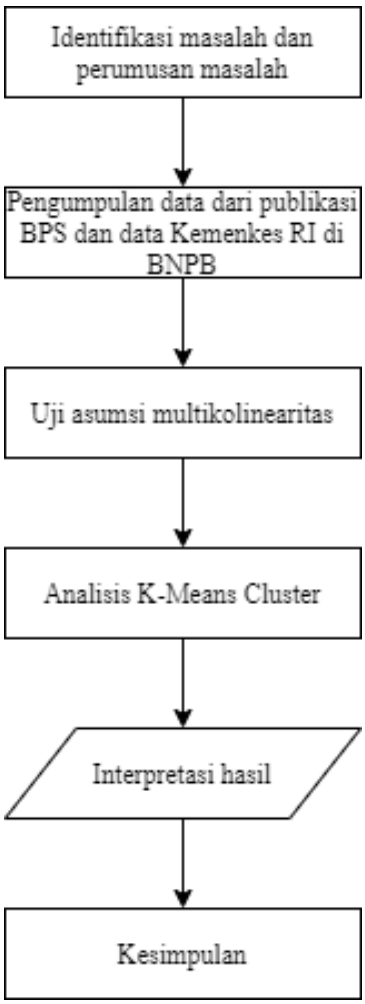
Penyebaran virus COVID-19 saat ini semakin luas, bahkan ke negara-negara lain yang terletak jauh, seperti Italia dan Iran. Hingga saat ini pun, vaksin untuk menyembuhkan COVID-19 masih dalam penelitian, dimana para peneliti terus berpacu dengan waktu dikarenakan makin banyaknya orang yang terinfeksi atau bahkan mengalami kematian diakibatkan keganasan virus ini.

Ada beberapa hal yang dapat dilakukan untuk menangani penyebaran COVID-19 yang makin meluas diantaranya dengan menganalisis laporan berita, *sentimen* masyarakat di media sosial, dan juga dokumen yang dirilis pemerintah tentang persebaran COVID-19. Dalam penelitian ini akan dianalisis kasus pengelompokkan *K-Means Cluster* Provinsi berdasarkan tingkat persebaran kasus COVID-19. Untuk mengetahui tingkat persebaran kasus COVID-19 kategori tinggi, sedang, dan rendah pada masing-masing Provinsi di Indonesia, ada beberapa aspek yang bisa diukur seperti jumlah penduduk, kepadatan penduduk, kasus positif terinfeksi COVID-19, pasien yang sembuh, dan pasien yang meninggal dunia. Berdasarkan tingkat persebaran kasus COVID-19 akan dilakukan clustering pada

Provinsi di Indonesia tahun 2020 dengan metode *K-Means Cluster*. Untuk melakukan analisis dengan metode *K-Means Cluster*, variabel-variabel tersebut harus melalui uji asumsi multikolinearitas terlebih dahulu dan setelah memenuhi uji asumsi multikolinearitas, dilanjutkan dengan analisis metode *K-Means Cluster*. Dalam penelitian ini akan dilakukan tahap-tahap untuk menganalisis data yang diperoleh antara lain.

1. Mengidentifikasi tingkat persebaran kasus COVID-19 dalam Provinsi di Indonesia.
2. Memperoleh data dari publikasi buku Badan Pusat Statistik di 34 Provinsi dan laporan data Kemenkes dari Badan Nasional Penanggulangan Bencana.
3. Menguji asumsi multikolinearitas.
4. Analisis dengan metode *K-Means Cluster*.
5. Menarik kesimpulan.

Berikut gambaran umum dari kerangka berpikir yang dapat dilihat pada gambar di bawah ini.



Gambar 2.1 Kerangka Berpikir

BAB III
METODE PENELITIAN

3.1 Populasi Penelitian

Populasi adalah wilayah generalisasi yang terdiri atas obyek/subyek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulanya (Sugiyono, 2011). Populasi dalam penelitian ini adalah 5 variabel yaitu jumlah penduduk; kepadatan penduduk; kasus positif terinfeksi COVID-19; pasien yang sembuh; dan pasien yang meninggal dunia.

3.2 Variabel Penelitian

Variabel penelitian pada dasarnya adalah segala sesuatu yang berbentuk apa saja yang ditetapkan oleh peneliti untuk dipelajari sehingga diperoleh informasi tentang hal tersebut, kemudian ditarik kesimpulannya. (Sugiyono, 2011) Adapun variabel yang digunakan dalam penelitian ini adalah:

1. Jumlah penduduk
2. Kepadatan penduduk
3. Kasus positif terinfeksi COVID-19
4. Pasien yang sembuh
5. Pasien yang meninggal dunia

3.3 Metode Pengumpulan Data

Menurut (Arikunto, 2010) metode dokumentasi adalah metode yang digunakan untuk mencari data mengenai hal-hal atau variable-variabel yang berupa catatan, transkip, buku, surat kabar, agenda, atau lain sebagainya. Dibandingkan dengan metode lain, maka metode ini agak tidak begitu sulit, dalam arti apabila ada kekeliruan sumber datanya masih tetap, belum berubah. Metode pengumpulan data yang digunakan dalam penyusunan Tugas Akhir ini adalah metode dokumentasi berupa data sekunder. Metode dokumentasi digunakan untuk memperoleh data dan informasi melalui pengumpulan data yang diperoleh dari publikasi buku Badan

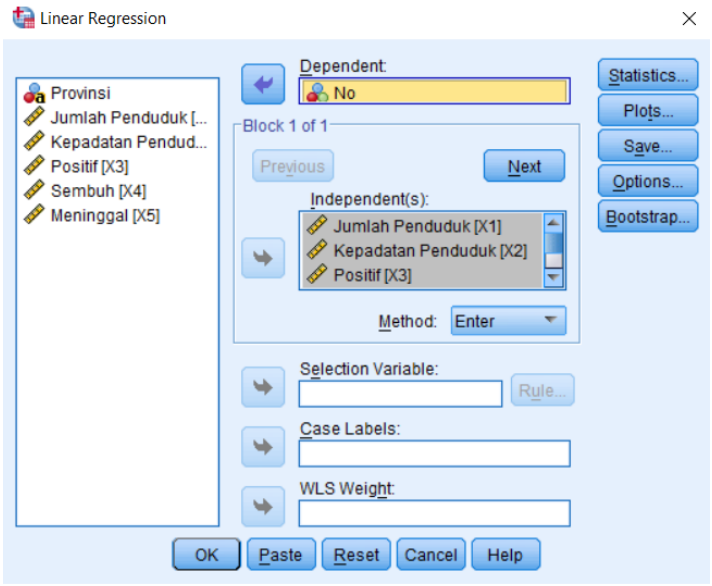
Pusat Statistik dan data Kemenkes RI di Badan Nasional Penanggulangan Bencana. Dalam hal ini data yang digunakan dari publikasi buku Badan Pusat Statistik adalah data berupa jumlah penduduk dan kepadatan penduduk yang dikumpulkan dari katalog yang berjudul Statistik Indonesia 2020 (BPS-Statistics Indonesia, 2020). Sedangkan data yang digunakan dari Kemenkes RI di Badan Nasional Penanggulangan Bencana adalah data Laporan Media Harian COVID-19 tanggal 31 Mei 2020 yang terdiri dari 3 variabel yaitu yaitu kasus positif terinfeksi COVID-19; pasien yang sembuh; dan pasien yang meninggal dunia (BNPB, 2020).

3.4 Analisis Data Dan Alur Penelitian

3.4.1 Analisis Data

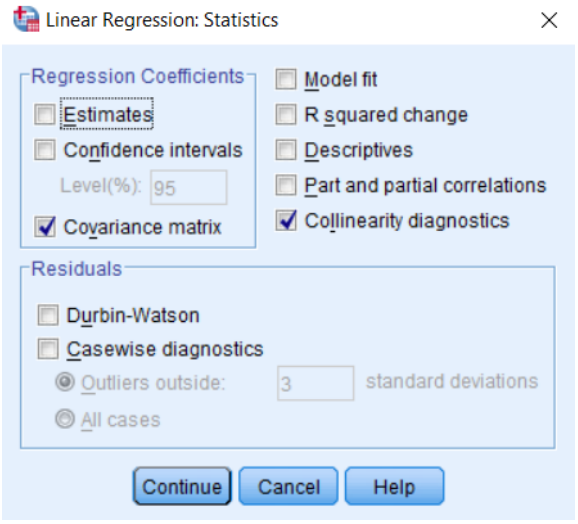
Langkah-langkah dalam analisis *K-Means Cluster* adalah sebagai berikut.

1. Pilih menu Analyse, Regression, Linier. Masukkan variabel No ke dependent dan masukkan variabel X_1, X_2, X_3, X_4 dan X_5 ke independent(s).



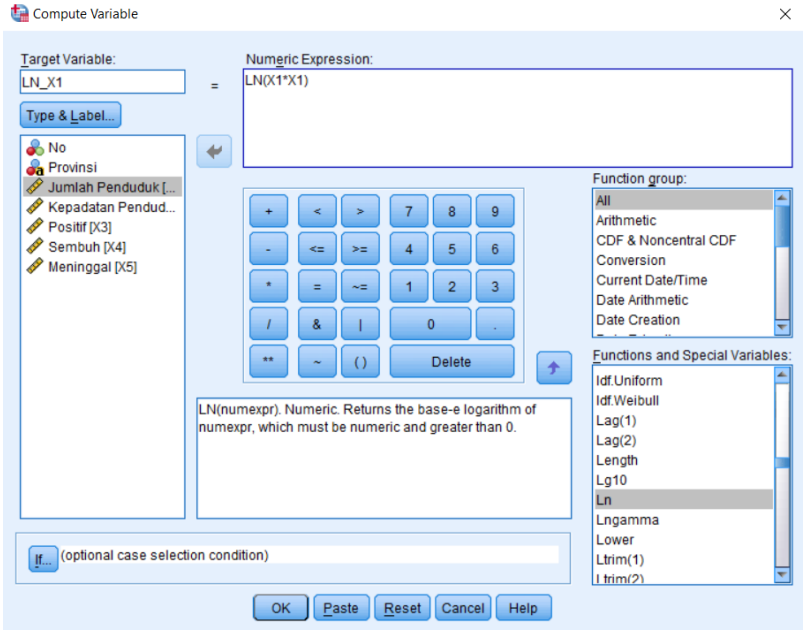
Gambar 3.1 Linear Regression

2. Pilih menu Statistics dan centang pilihan Covariance Matrix dan Collinearity Diagnostics kemudian klik continue dan ok.



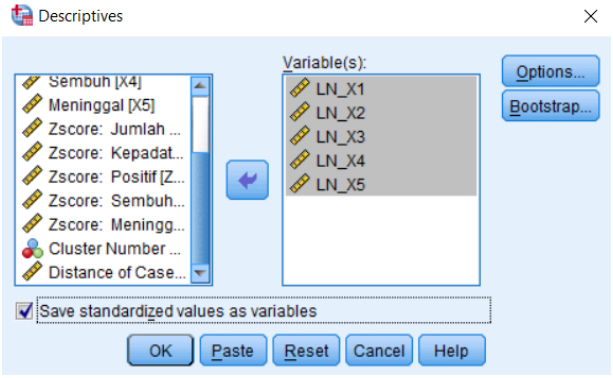
Gambar 3.2 Linear Regression Statistics

3. Jika data belum memenuhi uji asumsi multikolinearitas, maka data perlu ditransformasi ke Logaritma Normal terlebih dahulu. Pilih Transform, Compute Variable, isi Target Variable dengan nama baru sebagai variabel dari hasil transformasi dan isi pada kolom Numeric Expression sesuai gambar dan lakukan pada semua variabel.



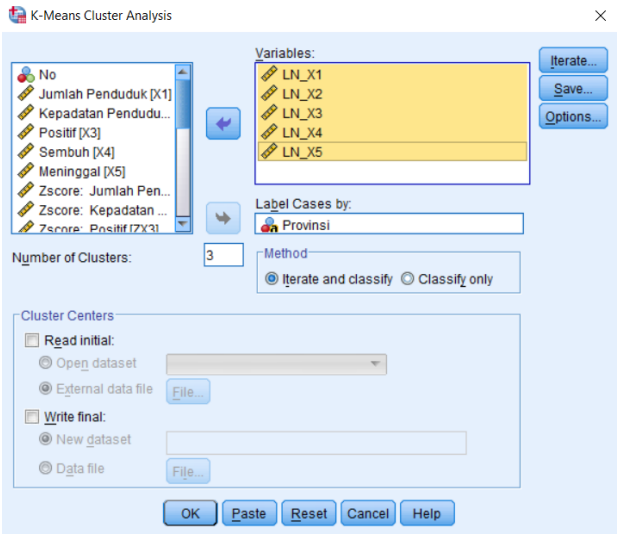
Gambar 3.3 Compute Variable

4. Pilih menu Analyze, kemudian pillih Descriptive Statistics dan Descriptives. Pada kotak dialog Descriptives, masukkan variabel LN_X_1, LN_X_2, LN_ X_3, LN_ X_4, dan LN_X_5 ke kotak variable(s) dan centang Save Standardized values as variables. Kemudian klik ok.



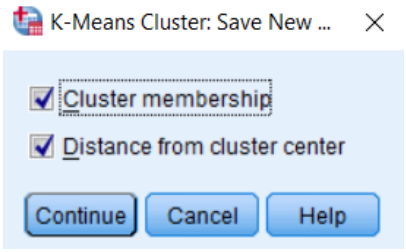
Gambar 3.4 Descriptives

5. Pilih menu Analyze, pilih Classify, kemudian pilih *K-Means Cluster*. Masukkan variabel Zscore: LN_X_1, LN_X_2, LN_ X_3, LN_ X_4, dan LN_X_5ke dalam variable(s) dan masukkan variabel Kab_Kota ke dalam Label Cases by. Kemudian isi kotak Number of Cluster dengan angka 3.



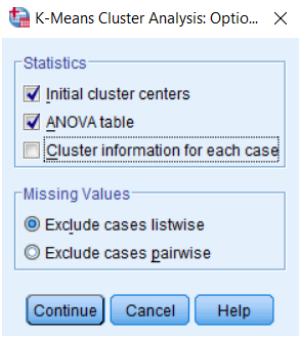
Gambar 3.5 *K-Means Cluster* Analysis

6. Pilih menu Save, kemudian centang kotak Cluster membership dan Distance from cluster center, pilih Continue.



Gambar 3.6 *K-Means Cluster*: Save

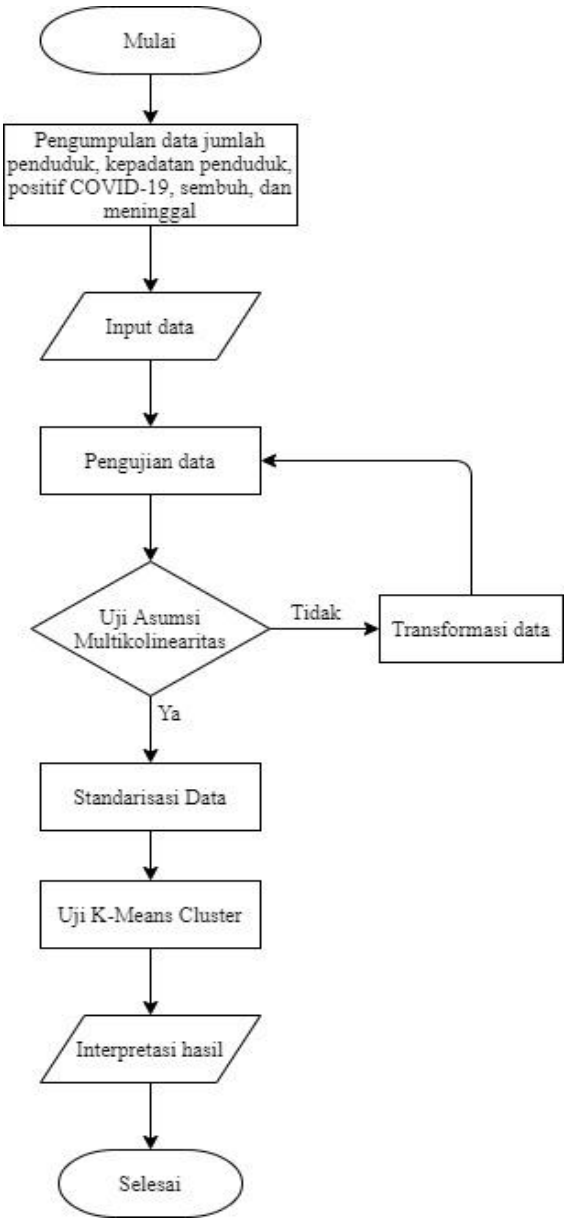
7. Pilih menu Options, kemudian centang kotak Initial cluster centers dan ANOVA table, pilih Continue.



Gambar 3.7 *K-Means Cluster*: Options

3.4.2 Alur Penelitian

Diagram analisis data:



Gambar 3.8 Langkah-langkah Analisis *K-Means Cluster*

BAB IV
HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

4.1.1 Data

Sebelum proses analisis, dilakukan pengumpulan semua data dari Badan Pusat Statistika dan Badan Nasional Penanggulangan Bencana. Setelah memperoleh semua data, maka data tersebut kemudian dianalisis menggunakan software SPSS 20 untuk analisis cluster yang akan mengelompokkan Provinsi di Indonesia berdasarkan tingkat kasus persebaran COVID-19 tanggal 31 Mei 2020 dengan variabel jumlah penduduk (X_1), kepadatan penduduk (X_2), positif (X_3), sembuh (X_4), dan meninggal (X_5).

4.1.2 Uji Multikolinearitas

Uji multikolinearitas dilakukan untuk menguji ada atau tidaknya variabel independen yang mempunyai kemiripan antar variabel independen lain (Yulianto & Hidayatullah, 2014). Jika data menunjukkan adanya multikolinearitas, maka salah satu cara yang dapat dilakukan yaitu melakukan Transformasi data ke dalam bentuk logaritma natural (Ghozali, 2018). Cara untuk menguji adanya multikolinieritas, yaitu:

1. Melihat nilai *tolerance*

   Terjadi multikolinieritas, jika nilai *tolerance* $\leq$ 0,1.

2. Melihat nilai VIF (*Variance Inflation Factor)*

   Terjadi multikolinieritas, jika nilai VIF $\geq$ 10.

Tabel 4.1 Hasil Uji Multikolinieritas

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Jumlah Penduduk	.434	2.304
	Kepadatan Penduduk	.085	11.813
	Positif	.008	129.797
	Sembuh	.033	30.713
	Meninggal	.013	76.156

Tabel di atas merupakan hasil pengujian dari multikolinieritas. Didapatkan bahwa variabel Jumlah Penduduk memiliki nilai *tolerance* lebih besar dari 0,1 serta nilai VIF nya kurang dari 10. Sedangkan pada variabel Kepadatan Penduduk, Positif, Sembuh, dan Meninggal memiliki nilai *tolerance* lebih kecil dari 0,1 serta nilai VIF nya lebih besar dari 10. Maka dapat disimpulkan bahwa pada semua variabel selain Jumlah Penduduk terjadi multikolinieritas. Karena data terjadi multikolinearitas, maka dilakukan Transformasi data ke dalam bentuk logaritma natural.

Tabel 4.2 Hasil Uji Multikolinieritas Dengan Transformasi Data

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	LN_X1	.852	1.174
	LN_X2	.870	1.149
	LN_X3	.347	2.884
	LN_X4	.368	2.721
	LN_X5	.441	2.267

a. Dependent Variable: No

Berdasarkan tabel di atas, didapatkan bahwa semua variabel memiliki nilai *tolerance* lebih besar dari 0,1 serta nilai VIF nya kurang dari 10. Jadi dapat disimpulkan bahwa semua variabel dengan data bentuk logaritma natural tidak terjadi multikolinearitas.

4.1.3 Analisis Deskriptif

Tabel 4.3 Statistik Deskriptif

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Jumlah Penduduk	34	743	49317	7884.58	11202.153
Kepadatan Penduduk	34	9.3	15900.0	741.059	2708.7090
Positif	34	20	7348	778.00	1470.436
Sembuh	34	14	2082	214.94	371.196
Meninggal	34	0	517	47.44	109.530
Valid N (listwise)	34				

Berdasarkan tabel yang didapat, terdapat 5 variabel yang dianalisis dari informasi N atau banyaknya data, minimum, maksimum, mean, dan standar deviasi. Pada variabel Jumlah Penduduk didapatkan data berjumlah 34 dengan nilai minimum 743, maksimum 49317, mean 7884,58, dan standar deviasi 11202,153. Pada variabel Kepadatan Penduduk didapatkan data berjumlah 34 dengan nilai minimum 9,3, maksimum 15900, mean 741,059, dan standar deviasi 2708,7090. Pada variabel Positif didapatkan data berjumlah 34 dengan nilai minimum 20, maksimum 7348, mean 778,00, dan standar deviasi 1470,436. Pada variabel Sembuh didapatkan data berjumlah 34 dengan nilai minimum 14, maksimum 2082, mean 214,94, dan standar deviasi 371,196. Pada variabel Meninggal didapatkan data berjumlah 34 dengan nilai minimum 0, maksimum 517, mean 47,44, dan standar deviasi 109,530.

4.1.4 Analisis Cluster dengan K-Means

Pada penelitian ini akan dilakukan pengujian analisis cluster dengan metode K-Means untuk clustering Provinsi di Indonesia berdasarkan tingkat kasus

persebaran COVID-19 tanggal 31 Mei 2020 menjadi 3 cluster, yaitu kasus persebaran tinggi, kasus persebaran sedang, dan kasus persebaran rendah.

Tabel 4.4 Initial Cluster Centers

Initial Cluster Centers

	Cluster		
	1	2	3
LN_X1	-2.87	-2.30	-5.90
LN_X2	3.44	-11.36	-2.86
LN_X3	2.99	-3.09	-3.95
LN_X4	3.23	-2.36	-6.87
LN_X5	2.91	-1.85	-4.20

Pada tabel 4.4 merupakan proses clustering data pertama sebelum data tersebut dilakukan iterasi dan data ini adalah proses untuk pembentukan 3 cluster. Jadi, tabel di atas tidak dianalisis.

Tabel 4.5 Iteration History

Iteration History[a]

Iteration	Change in Cluster Centers		
	1	2	3
1	4.946	5.432	6.270
2	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 2. The minimum distance between initial centers is 10,573.

Tabel di atas merupakan proses iterasi dalam pengelompokan cluster dari tabel initial dan menghasilkan proses iterasi sebanyak 2 kali. Pada iterasi 1 terjadi centeroid yang tidak signifikan dan pada iterasi 2 terjadi centeroid yang signifikan.

Jadi, semua cluster sudah terbentuk dan iterasi berhenti pada iterasi 2 dengan jarak minimum 10,573.

Tabel 4.6 Final Cluster Centers

Final Cluster Centers

	Cluster		
	1	2	3
LN_X1	-.12	.51	-2.25
LN_X2	.30	-7.48	-2.92
LN_X3	1.50	-.92	-2.41
LN_X4	1.70	-1.37	-2.37
LN_X5	1.33	-.90	-2.37

Pada tabel di atas adalah hasil dari proses akhir dalam clustering yang membentuk cluster sebanyak 3 untuk masing-masing variabel. Variabel pada tabel Final Cluster Centers merupakan hasil untuk nilai standarisasi. Angka-angka yang terbentuk di atas mempunyai arti. Angka negatif memiliki arti bahwa data berada di bawah rata-rata total dan angka positif memiliki arti bahwa data berada di atas rata-rata total. Untuk mengetahui pengaruh pada variabel di atas dilakukan perhitungan dengan rumus:

$$X = \mu + z.\sigma$$

Keterangan:

X = rata-rata data atau sampel

μ = rata-rata populasi

σ = standar deviasi

z = nilai standarisasi

Berikut adalah cluster dengan tafsiran angka untuk masing-masing variabel beserta tabelnya:

1. Variabel Jumlah Penduduk

a. Cluster 1 : rata-rata variabel jumlah penduduk + (nilai z cluster 1 × standar deviasi rata-rata variabel jumlah penduduk) = $7884{,}58 + (-0{,}12 \times 11202{,}153) = 6485{,}78$.

b. Cluster 2 : rata-rata variabel jumlah penduduk + (nilai z cluster 2 × standar deviasi rata-rata variabel jumlah penduduk) = $7884{,}58 + (-0{,}51 \times 11202{,}153) = 13616{,}74$.

c. Cluster 3 : rata-rata variabel jumlah penduduk + (nilai z cluster 3 × standar deviasi rata-rata variabel jumlah penduduk) = $7884{,}58 + (-2{,}25 \times 11202{,}153) = -17346{,}6$.

2. Variabel Kepadatan Penduduk

a. Cluster 1 : rata-rata variabel kepadatan penduduk + (nilai z cluster 1 × standar deviasi rata-rata variabel jumlah penduduk) = $741{,}059 + (0{,}30 \times 2708{,}7090) = 1551{,}988$.

b. Cluster 2 : rata-rata variabel kepadatan penduduk + (nilai z cluster 2 × standar deviasi rata-rata variabel jumlah penduduk) = $741{,}059 + (-7{,}48 \times 2708{,}7090) = -19533{,}6$.

c. Cluster 3 : rata-rata variabel kepadatan penduduk + (nilai z cluster 3 × standar deviasi rata-rata variabel jumlah penduduk) = $741{,}059 + (-2{,}92 \times 2708{,}7090) = -7181{,}89$.

3. Variabel Positif

a. Cluster 1 : rata-rata variabel positif + (nilai z cluster 1 × standar deviasi rata-rata variabel jumlah penduduk) = $778 + (1{,}50 \times 1470{,}436) = 2990{,}695$.

b. Cluster 2 : rata-rata variabel positif + (nilai z cluster 2 × standar deviasi rata-rata variabel jumlah penduduk) = $778 + (-0{,}92 \times 1470{,}436) = -577{,}136$.

c. Cluster 3 : rata-rata variabel positif + (nilai z cluster 3 × standar deviasi rata-rata variabel jumlah penduduk) = $778 + (-2{,}41 \times 1470{,}436) = -2765{,}03$.

4. Variabel Sembuh

a. Cluster 1 : rata-rata variabel sembuh + (nilai z cluster 1 × standar deviasi rata-rata variabel jumlah penduduk) = $214{,}94 + (1{,}70 \times 371{,}196) = 846{,}056$.

b. Cluster 2 : rata-rata variabel sembuh + (nilai z cluster 2 × standar deviasi rata-rata variabel jumlah penduduk) = $214{,}94 + (-1{,}37 \times 371{,}196) = -293{,}093$.

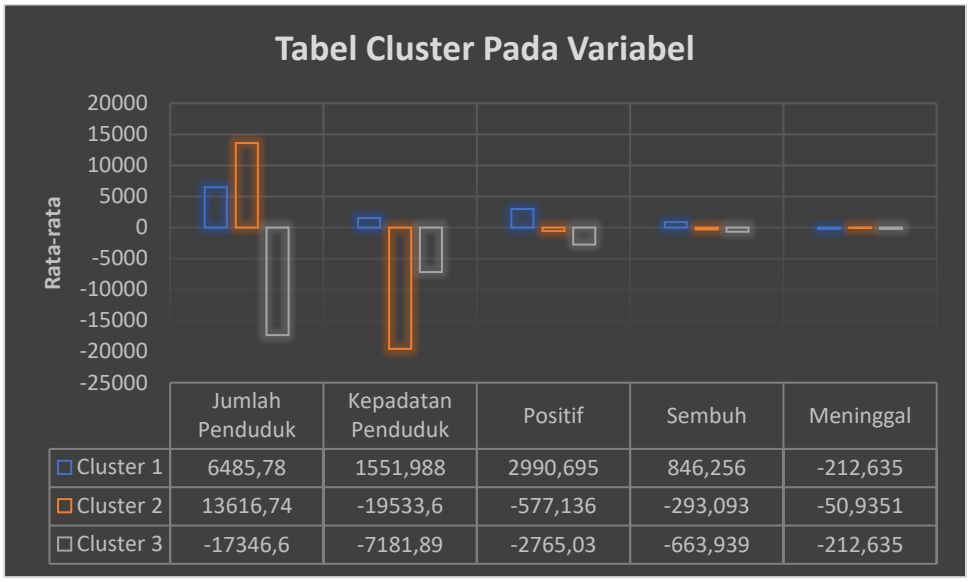
c. Cluster 3 : rata-rata variabel sembuh + (nilai z cluster 3 × standar deviasi rata-rata variabel jumlah penduduk) = $214{,}94 + (-2{,}37 \times 371{,}196) = -663{,}939$.

5. Variabel Meninggal

a. Cluster 1 : rata-rata variabel meninggal + (nilai z cluster 1 × standar deviasi rata-rata variabel jumlah penduduk) = $47{,}44 + (1{,}33 \times 109{,}530) = 193{,}0667$.

b. Cluster 2 : rata-rata variabel meninggal + (nilai z cluster 2 × standar deviasi rata-rata variabel jumlah penduduk) = $47{,}44 + (-0{,}90 \times 109{,}530) = -50{,}9351$.

c. Cluster 3 : rata-rata variabel meninggal + (nilai z cluster 3 × standar deviasi rata-rata variabel jumlah penduduk) = $47{,}44 + (-2{,}37 \times 109{,}530) = -212{,}635$.



	Jumlah Penduduk	Kepadatan Penduduk	Positif	Sembuh	Meninggal
Cluster 1	6485,78	1551,988	2990,695	846,256	-212,635
Cluster 2	13616,74	-19533,6	-577,136	-293,093	-50,9351
Cluster 3	-17346,6	-7181,89	-2765,03	-663,939	-212,635

Gambar 4.1 Rata-Rata Cluster Pada Masing-Masing Variabel

Berdasarkan perhitungan yang dihasilkan dari masing-masing variabel, didapatkan karakteristik pada masing-masing cluster. Berikut adalah penjelasan untuk setiap cluster:

1. Cluster 1

   Karakteristik yang terdapat pada cluster 1 berisi variabel jumlah penduduk yang berada di bawah rata-rata populasi, variabel kepadatan penduduk berada di atas rata-rata populasi, variabel positif berada di atas rata-rata populasi, variabel sembuh berada di atas rata-rata populasi, dan variabel meninggal berada di atas rata-rata populasi.
2. Cluster 2

   Karakteristik yang terdapat pada cluster 2 berisi variabel jumlah penduduk yang berada di atas rata-rata populasi, variabel kepadatan penduduk berada di bawah rata-rata populasi, variabel positif berada di bawah rata-rata populasi, variabel sembuh berada di bawah rata-rata populasi, dan variabel meninggal berada di bawah rata-rata populasi.

3. Cluster 3

   Karakteristik yang terdapat pada cluster 3 berisi variabel jumlah penduduk yang berada di bawah rata-rata populasi, variabel kepadatan penduduk berada di bawah rata-rata populasi, variabel positif berada di bawah rata-rata populasi, variabel sembuh berada di bawah rata-rata populasi, dan variabel meninggal berada di bawah rata-rata populasi.

4.1.5 Melihat Tingkat Signifikansi Antar Cluster

Untuk melihat pengujian tingkat signifikansi antar cluster dan mengetahui perbedaan di setiap cluster, maka perlu dilakukan uji ANOVA. Ketentuan penggunaan angka F dalam analisis cluster ialah bahwa semakin besar angka F hitung (jika dilakukan uji Hipotesis, maka F hitung akan lebih besar dari F tabel) dan tingkat signifikansi (sig) < 0,05; maka semakin besar perbedaan antara ketiga cluster yang terbentuk (Bastian et al., 2018). Berikut adalah hasil dari uji ANOVA:

Tabel 4.7 ANOVA

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
LN_X1	13.671	2	2.815	31	4.857	.015
LN_X2	40.818	2	1.532	31	26.641	.000
LN_X3	16.421	2	2.096	31	7.835	.002
LN_X4	16.222	2	2.157	31	7.522	.002
LN_X5	14.929	2	1.553	31	9.611	.001

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Berdasarkan tabel yang didapat, pada kolom Cluster adalah besaran *between cluster means* dan kolom Error adalah besaran *within cluster means.* Sementara pada kolom F didapat dari rumus sebagai berikut.

$$F = \frac{between\ cluster\ mean}{within\ cluster\ mean}$$

Hipotesis:

H_0: Ketiga cluster tidak mempunyai perbedaan yang signifikan.

H_1: Ketiga cluster mempunyai perbedaan yang signifikan.

Jika angka signifikan > 0.05; H_0 diterima H_1 ditolak.

Jika angka signifikan < 0,05; H_0 ditolak H_1 diterima.

Pada tabel 4.8 dapat dilihat variabel yang membedakan berturut-turut:

1. Jumlah Penduduk dengan nilai F = 4.857 dan sig. = 0.015.
2. Kepadatan Penduduk dengan nilai F = 26.641 dan sig. = 0.000.
3. Positif dengan nilai F = 7.835 dan sig. = 0.002.
4. Sembuh dengan nilai F = 7.522 dan sig. = 0.002.
5. Meninggal dengan nilai F = 9.611 dan sig. = 0.001.

Berdasarkan hasil yang didapatkan bahwa semua variabel memiliki nilai signifikansi (sig.) < 0,05, maka ketiga cluster mempunyai perbedaan yang signifikan dan diperoleh hasil bahwa variabel kepadatan penduduk adalah variabel yang paling membedakan anggota dari ketiga cluster karena mempunyai nilai F terbesar diantara variabel yang lain yaitu 26,641 dan nilai signifikan 0,000.

4.1.6 Jumlah Anggota Di Masing-Masing Cluster

Setelah dilakukan uji ANOVA, didapatkan tabel untuk melihat jumlah anggota pada setiap cluster yang terbentuk yang bisa dilihat sebagai berikut.

Tabel 4.8 Number of Cases in Each Cluster

Number of Cases in each Cluster

Cluster	1	2.000
	2	3.000
	3	29.000
Valid		34.000
Missing		.000

Berdasarkan tabel di atas, tidak ada data yang missing dan pada cluster 1 terdapat jumlah anggota sebanyak 2 provinsi, cluster 2 sebanyak 3 provinsi, dan cluster 3 sebanyak 29 provinsi.

4.1.7 Profiling Komposisi Cluster

Setelah mengetahui jumlah anggota untuk masing-masing cluster, didapatkan tabel cluster membership yang bisa dilihat di Lampiran 1 **Hasil Ouput SPSS Analisis *K-Means Cluster*** pada kelompok cluster.

Penafsiran dari tabel cluster membership adalah sebagai berikut.

1. Cluster 1

   Pada cluster 1 menghasilkan 2 provinsi, yaitu Provinsi DKI Jakarta yang memliki jarak sebesar 4,946 dan Provinsi Jawa Barat memiliki jarak sebesar 4,946. Meskipun memiliki jarak antar titik yang sama, tetapi Provinsi DKI Jakarta yang lebih dominan dengan jumlah penduduk sebanyak 10558, kepadatan penduduk sebanyak 15900, positif sebanyak 7348, sembuh sebanyak 2082, dan meninggal sebanyak 517. Karena kepadatan penduduknya melebihi jumlah penduduk, maka daerah tersebut memiliki daerah yang padat sehingga penyebaran pada COVID-19 menjadi lebih mudah dan mempunyai kasus penyakit COVID-19 yang lebih tinggi.

2. Cluster 2

   Pada cluster 2 menghasilkan 3 provinsi, yaitu Provinsi Bali memiliki jarak sebesar 5,432, Provinsi Jawa Tengah memiliki jarak sebesar 4,211, dan Provinsi Jawa Timur memiliki jarak sebesar 4,925. Dari ketiga provinsi tersebut, Provinsi Jawa Timur yang memiliki kasus COVID-19 paling dominan dari positif sebanyak 4857, sembuh sebanyak 609, dan meninggal sebanyak 396 dengan jumlah penduduk sebanyak 39699 dan kepadatan penduduk sebanyak 831.

3. Cluster 3

   Pada cluster 3 menghasilkan 29 provinsi, yaitu Provinsi Aceh memiliki jarak sebesar 1,841, Provinsi Banten memiliki jarak sebesar 4,449, Provinsi Bangka Belitung memiliki jarak sebesar 1,939, Provinsi Bengkulu memiliki jarak sebesar 1,796, Provinsi DI Yogyakarta memiliki jarak sebesar 1,673, Provinsi Jambi memiliki jarak sebesar 1,626, Provinsi Kalimantan Barat memiliki jarak sebesar 1,196, Provinsi Kalimantan Timur memiliki jarak sebesar 2,070, Provinsi Kalimantan Tengah memiliki jarak sebesar 1,840, Provinsi Kalimantan Selatan memiliki jarak sebesar 2,281, Provinsi Kalimantan Utara memiliki jarak sebesar 1,716, Provinsi Kepulauan Riau memiliki jarak sebesar 1,197, Provinsi Nusa Tenggara Barat memiliki jarak sebesar 2,517, Provinsi Sumatera Selatan memiliki jarak sebesar 6,270, Provinsi Sumatera Barat memiliki jarak sebesar 2,475, Provinsi Sulawesi Utara memiliki jarak sebesar 2,572, Provinsi Sumatera Utara memiliki jarak sebesar 3,679, Provinsi Sulawesi Tenggara memiliki jarak sebesar 1,051, Provinsi Sulawesi Selatan memiliki jarak sebesar 3,789, Provinsi Sulawesi Tengah memiliki jarak sebesar 1,269, Provinsi Lampung memiliki jarak sebesar 3,869, Provinsi Riau memiliki

jarak sebesar 2,921, Provinsi Maluku Utara memiliki jarak sebesar 1,740, Provinsi Maluku memiliki jarak sebesar 1,504, Provinsi Papua Barat memiliki jarak sebesar 1,740, Provinsi Papua memiliki jarak sebesar 3,030, Provinsi Sulawesi Barat memiliki jarak sebesar 1,804, Provinsi Nusa Tenggara Timur memiliki jarak sebesar 1,779, dan Provinsi Gorontalo memiliki jarak sebesar 1,902. Dari 29 provinsi tersebut, Provinsi Sulawesi Selatan yang memiliki kasus COVID-19 paling dominan dari positif sebanyak 1541, sembuh sebanyak 596, dan meninggal sebanyak 74 dengan jumlah penduduk sebanyak 8851 dan kepadatan penduduk sebanyak 193.

Berikut adalah tabel berisi Provinsi pada masing-masing cluster dan gambar berisi Provinsi yang memiliki kasus tertinggi pada setiap cluster.

Tabel 4.9 Provinsi Pada Setiap Cluster

	Cluster 1	Cluster 2	Cluster 3
Provinsi	DKI Jakarta dan Jawa Barat	Bali, Jawa Tengah, Jawa Timur	Aceh, Banten, Bangka Belitung, Bengkulu, DI Yogyakarta, Jambi, Kalimantan Barat, Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Kepulauan Riau, Nusa Tenggara Barat, Sumatera Selatan, Sumatera Barat, Sulawesi Utara, Sumatera Utara, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Lampung, Riau, Maluku Utara, Maluku, Papua Barat, Papua, Sulawesi Barat, Nusa Tenggara Timur, dan Gorontalo

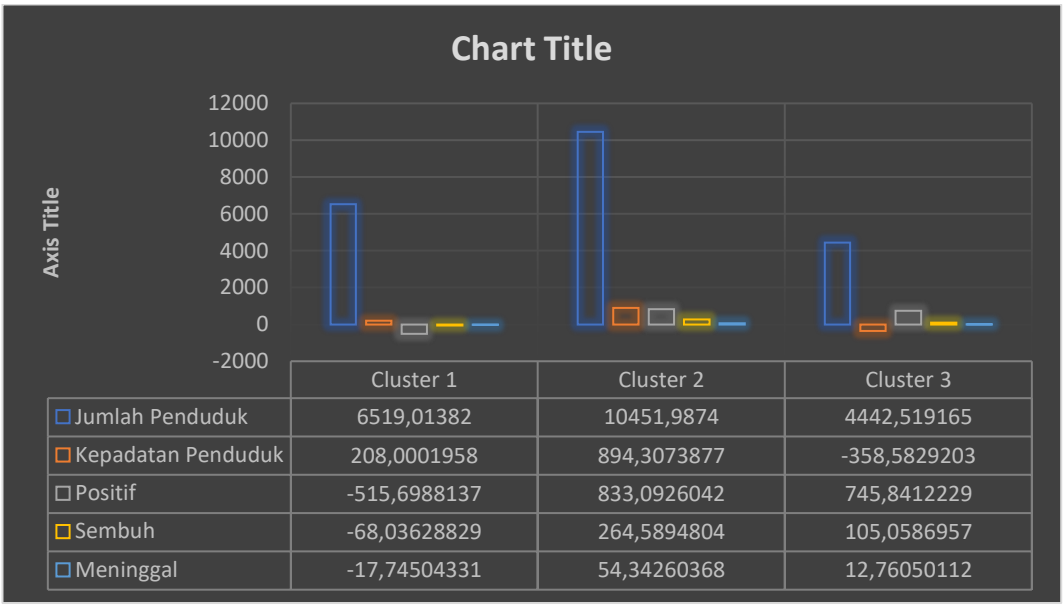
Gambar 4.2 Provinsi Yang Memiliki Kasus Tertinggi Pada Setiap Cluster

Pada cluster 3 yang berisi 29 Provinsi akan dipecah kembali menjadi 3 cluster yang terbagi menjadi kasus tingkat pertama, kedua, dan ketiga dari cluster yang memiliki kategori persebaran kasus COVID-19 yang rendah. Berikut adalah tabel statistik deskriptif dan gambar hasil tafsiran angka untuk masing-masing variabel pada cluster 3.

Tabel 4.10 Statistik Deskriptif Pada Cluster 3

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Jumlah Penduduk	29	743	14563	4463.76	3449.558
Kepadatan Penduduk	29	9.3	1338.0	181.459	311.3429
Positif	29	20	1541	348.93	354.817
Sembuh	29	14	596	114.07	121.459
Meninggal	29	0	80	16.62	22.873
Valid N (listwise)	29				



	Cluster 1	Cluster 2	Cluster 3
Jumlah Penduduk	6519,01382	10451,9874	4442,519165
Kepadatan Penduduk	208,0001958	894,3073877	-358,5829203
Positif	-515,6988137	833,0926042	745,8412229
Sembuh	-68,03628829	264,5894804	105,0586957
Meninggal	-17,74504331	54,34260368	12,76050112

Gambar 4.3 Hasil Tafsiran Angka Untuk Masing-Masing Variabel Pada Cluster 3

Berdasarkan perhitungan yang dihasilkan dari masing-masing variabel pada cluster 3, didapatkan karakteristiknya sebagai berikut:

1. Cluster 1

   Karakteristik yang terdapat pada cluster 1 berisi variabel jumlah penduduk dan kepadatan penduduk berada di atas rata-rata populasi. Sedangkan variabel positif, sembuh, dan meninggal berada di bawah rata-rata populasi. Provinsi yang berada pada cluster 1 termasuk dalam kategori tingkat ketiga.

2. Cluster 2

   Karakteristik yang terdapat pada cluster 2 berisi variabel jumlah penduduk, kepadatan penduduk, positif, sembuh, dan meninggal berada di atas rata-rata populasi. Provinsi yang berada pada cluster 2 termasuk dalam kategori tingkat pertama.

3. Cluster 3

   Karakteristik yang terdapat pada cluster 3 berisi variabel jumlah penduduk dan positif berada di atas rata-rata populasi. Sedangkan variabel kepadatan penduduk, sembuh, dan meninggal berada di bawah

rata-rata populasi. Provinsi yang berada pada cluster 3 termasuk dalam kategori tingkat kedua.

4.1.8 Pengelompokkan Provinsi Pada Cluster 3

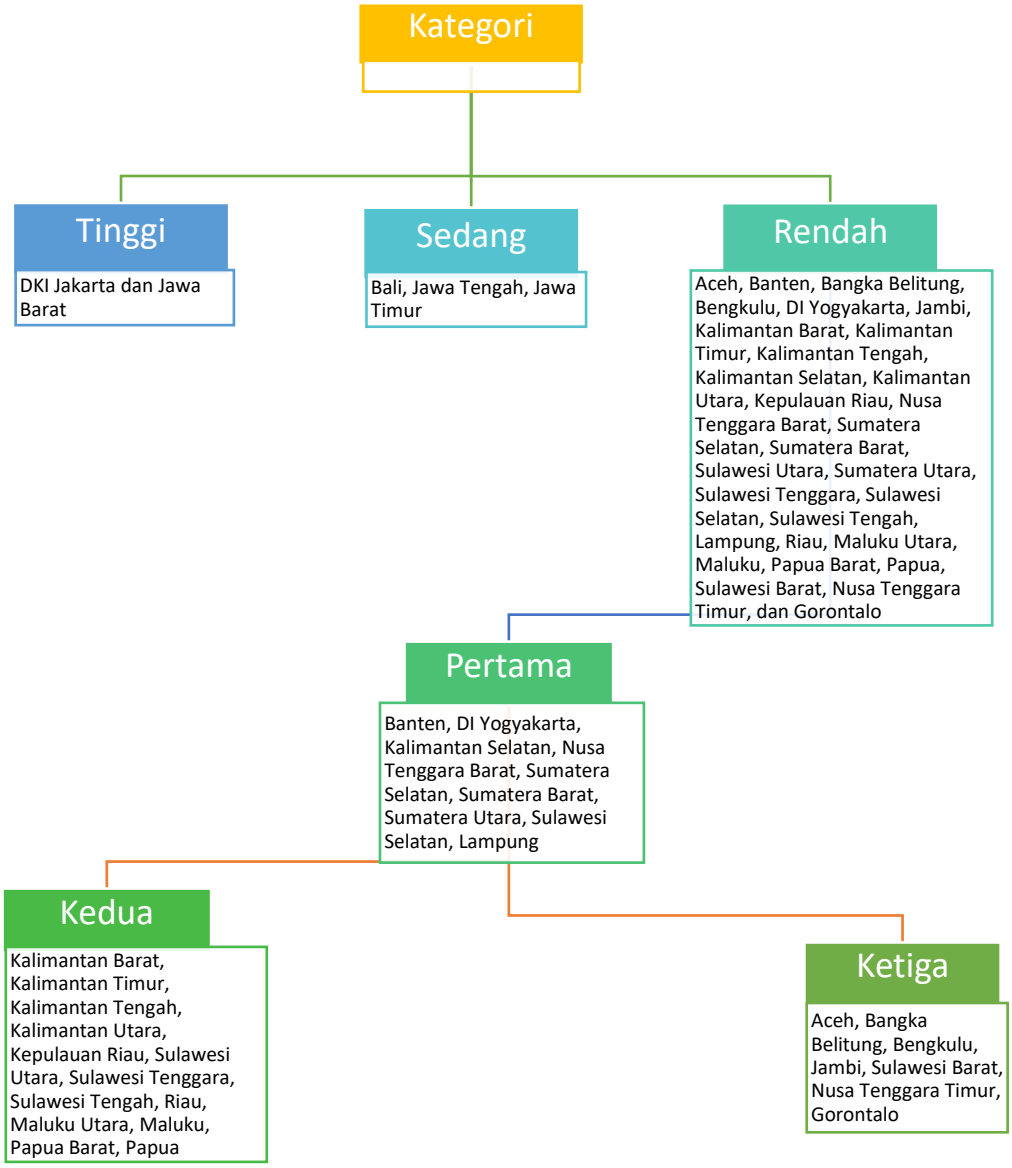
Setelah dilakukan uji *K-Means Cluster* pada cluster 3 yang berisi 29 Provinsi, didapatkan tabel cluster membership yang dapat dilihat outputnya pada Lampiran 2 **Hasil Ouput SPSS Analisis *K-Means Cluster* Pada Cluster 3** dan disimpulkan sebagai berikut.

Tabel 4.11 Provinsi Pada Setiap Cluster

Cluster 1	Cluster 2	Cluster 3
• Aceh, Bangka Belitung, Bengkulu, Jambi, Sulawesi Barat, Nusa Tenggara Timur, Gorontalo	• Banten, DI Yogyakarta, Kalimantan Selatan, Nusa Tenggara Barat, Sumatera Selatan, Sumatera Barat, Sumatera Utara, Sulawesi Selatan, Lampung	• Kalimantan Barat, Kalimantan Timur, Kalimantan Tengah, Kalimantan Utara, Kepulauan Riau, Sulawesi Utara, Sulawesi Tenggara, Sulawesi Tengah, Riau, Maluku Utara, Maluku, Papua Barat, Papua

Berdasarkan tabel di atas, pada tingkat pertama berisi Provinsi yang berada di cluster 2, yaitu Provinsi Banten, DI Yogyakarta, Kalimantan Selatan, Nusa Tenggara Barat, Sumatera Selatan, Sumatera Barat, Sumatera Utara, Sulawesi Selatan, dan Lampung. Pada tingkat kedua berisi Provinsi yang berada di cluster 3, yaitu Provinsi Kalimantan Barat, Kalimantan Timur, Kalimantan Tengah, Kalimantan Utara, Kepulauan Riau, Sulawesi Utara, Sulawesi Tenggara, Sulawesi Tengah, Riau, Maluku Utara, Maluku, Papua Barat, dan Papua. Sedangkan pada tingkat ketiga berisi Provinsi yang berada di cluster 1, yaitu Provinsi Aceh, Bangka Belitung, Bengkulu, Jambi, Sulawesi Barat, Nusa Tenggara Timur, dan Gorontalo.

4.1.9 Diagram Hasil Analisis K-Means Cluster



Gambar 4.4 Diagram Hasil Analisis K-Means Cluster

4.2 Pembahasan

Dalam hasil analisis ini, terdapat 5 variabel untuk mengelompokkan Provinsi di Indonesia berdasarkan tingkat kasus persebaran COVID-19 pada tanggal 31 Mei 2020, yaitu variabel jumlah penduduk (X_1), kepadatan penduduk (X_2), positif (X_3), sembuh (X_4), dan meninggal (X_5). Sebelum dilakukan uji *K-Means Cluster*, data harus memenuhi asumsi klasik dalam clustering yaitu uji multikolinearitas yang berfungsi untuk menguji ada atau tidaknya variabel independen yang mempunyai kemiripan antar variabel independen lain. Karena data terjadi multikolinearitas, maka data tersebut dilakukan standarisasi dengan melakukan transformasi data ke dalam bentuk logaritma natural. Setalah dilakukan transformasi data dan dilakukan ulang uji multikolinearitas, data tersebut tidak terjadi multikolinearitas sehingga data dapat dilanjutkan uji *K-Means Cluster*.

Dalam hasil analisis *K-Means Cluster* dapat disimpulkan kelompok cluster mana yang memiliki tingkat kasus persebaran COVID-19 yang tinggi, sedang, dan rendah berdasarkan karakteristik pada masing-masing cluster. Pada cluster dengan tafsiran angka untuk masing-masing variabel berfungsi untuk menentukan cluster mana yang memiliki tingkat tinggi, sedang, dan rendah.

Pada cluster 1 berdasarkan tafsiran angka yang didapat dapat disimpulkan bahwa kepadatan penduduk, positif, sembuh, dan meninggal memiliki nilai rata-rata yang lebih tinggi dibandingkan dengan cluster 2 dan 3. Sedangkan jumlah penduduk lebih rendah dari cluster 2 dan lebih tinggi dari cluster 3. Maka dapat disimpulkan bahwa cluster 1 termasuk dalam Provinsi yang memiliki tingkat kasus persebaran COVID-19 yang tinggi. Provinsi yang masuk dalam cluster ini adalah Provinsi DKI Jakarta dan Provinsi Jawa Barat.

Pada cluster 2 berdasarkan tafsiran angka yang didapat dapat disimpulkan bahwa jumlah penduduk memiliki nilai rata-rata yang lebih tinggi dan kepadatan penduduk yang lebih rendah dibandingkan dengan cluster 1 dan 3. Sedangkan positif, sembuh, dan meninggal memiliki rata-rata lebih rendah dari cluster 1 dan

lebih tinggi dari cluster 3. Maka dapat disimpulkan bahwa cluster 2 termasuk dalam Provinsi yang memiliki tingkat kasus persebaran COVID-19 yang sedang. Provinsi yang masuk dalam cluster ini adalah Provinsi Bali, Provinsi Jawa Tengah, dan Provinsi Jawa Timur.

Pada cluster 3 berdasarkan tafsiran angka yang didapat dapat disimpulkan bahwa jumlah penduduk memiliki nilai rata-rata yang lebih rendah dibandingkan dengan cluster 1 dan 2. Sedangkan kepadatan penduduk memiliki rata-rata lebih rendah dari cluster 1 dan lebih tinggi dari cluster 2. Untuk positif, sembuh, dan meninggal memiliki rata-rata lebih rendah dari cluster 1 dan 2. Maka dapat disimpulkan bahwa cluster 3 termasuk dalam Provinsi yang memiliki tingkat kasus persebaran COVID-19 yang rendah. Provinsi yang masuk dalam cluster ini adalah Provinsi Aceh, Provinsi Banten, Provinsi Bangka Belitung, Provinsi Bengkulu, Provinsi DI Yogyakarta, Provinsi Jambi, Provinsi Kalimantan Barat, Provinsi Kalimantan Timur, Provinsi Kalimantan Tengah, Provinsi Kalimantan Selatan, Provinsi Kalimantan Utara, Provinsi Kepulauan Riau, Provinsi Nusa Tenggara Barat, Provinsi Sumatera Selatan, Provinsi Sumatera Barat, Provinsi Sulawesi Utara, Provinsi Sumatera Utara, Provinsi Sulawesi Tenggara, Provinsi Sulawesi Selatan, Provinsi Sulawesi Tengah, Provinsi Lampung, Provinsi Riau, Provinsi Maluku Utara, Provinsi Maluku, Provinsi Papua Barat, Provinsi Papua, Provinsi Sulawesi Barat, Provinsi Nusa Tenggara Timur, dan Provinsi Gorontalo.

Hasil cluster yang didapat berdasarkan uji ANOVA menunjukkan bahwa semua variabel yang digunakan mempunyai perbedaan yang signifikan pada ketiga cluster yang terbentuk dan variabel yang memiliki perbedaan paling besar pada ketiga cluster yang terbentuk adalah variabel kepadatan penduduk dengan nilai F sebesar 26,641 dan nilai signifikan 0,000. Karena variabel kepadatan penduduk memiliki perbedaan paling besar terhadap cluster yang terbentuk dalam tingkat kasus persebaran COVID-19, maka Provinsi yang memiliki jumlah kepadatan penduduk terbesar di Indonesia yaitu Provinsi DKI Jakarta dan DKI Jakarta

termasuk dalam cluster 1 yang memiliki tingkat kasus persebaran COVID-19 yang tinggi. Jumlah anggota pada setiap cluster yang terbentuk berdasarkan tingkat kasus persebaran COVID-19 mengelompokkan 34 provinsi menjadi 3 cluster. Pada cluster 1 terdapat jumlah anggota sebanyak 2 provinsi, cluster 2 sebanyak 3 provinsi, dan cluster 3 sebanyak 29 provinsi dengan tidak adanya variabel yang missing (hilang). Sehingga dapat disimpulkan bahwa semua kasus berdasarkan provinsi di Indonesia lengkap dan masuk dalam ketiga cluster.

Pada cluster 3 yang berisi 29 Provinsi dibagi lagi menjadi 3 cluster dengan mengelompokkan berdasarkan tingakatannya. Didapatkan dengan uji *K-Means Cluster* pada cluster 3 didapatkan pada tingkat pertama diisi dengan cluster 2 yang berisi 9 Provinsi, yaitu Provinsi Banten, DI Yogyakarta, Kalimantan Selatan, Nusa Tenggara Barat, Sumatera Selatan, Sumatera Barat, Sumatera Utara, Sulawesi Selatan, dan Lampung. Pada tingkat kedua diisi dengan cluster 3 yang berisi 13 Provinsi, yaitu Provinsi Kalimantan Barat, Kalimantan Timur, Kalimantan Tengah, Kalimantan Utara, Kepulauan Riau, Sulawesi Utara, Sulawesi Tenggara, Sulawesi Tengah, Riau, Maluku Utara, Maluku, Papua Barat, dan Papua. Sedangkan pada tingkat ketiga diisi dengan cluster 1 yang berisi 7 Provinsi, yaitu Provinsi Aceh, Bangka Belitung, Bengkulu, Jambi, Sulawesi Barat, Nusa Tenggara Timur, dan Gorontalo.

BAB V
PENUTUP

5.1 Simpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, maka dapat disimpulkan sebagai berikut:

1. Penerapan dengan metode *K-Means Cluster* digunakan untuk mengelompokkan 34 Provinsi di Indonesia berdasarkan tingkat persebaran kasus COVID-19 pada tanggal 31 Mei 2020. Sebelum analisis *K-Means Cluster*, dilakukan uji asumsi multikolinieritas terlebih dahulu. Setelah memenuhi asumsi, dilanjutkan analisis *K-Means Cluster*. Analisis dengan metode *K-Means Cluster* dalam pengelompokkan 34 Provinsi di Indonesia berdasarkan tingkat persebaran kasus COVID-19 dibentuk menjadi 5 cluster.

a. Cluster 1 termasuk dalam Provinsi yang memiliki tingkat kasus persebaran COVID-19 yang tinggi. Provinsi yang masuk dalam cluster ini adalah Provinsi DKI Jakarta dan Provinsi Jawa Barat.
b. Cluster 2 termasuk dalam provinsi yang memiliki tingkat kasus persebaran COVID-19 yang sedang. Provinsi yang masuk dalam cluster ini adalah Provinsi Bali, Provinsi Jawa Tengah, dan Provinsi Jawa Timur.
c. Cluster 3 termasuk dalam Provinsi yang memiliki tingkat kasus persebaran COVID-19 yang rendah tingkat pertama. Provinsi yang masuk dalam cluster ini adalah Provinsi Banten, DI Yogyakarta, Kalimantan Selatan, Nusa Tenggara Barat, Sumatera Selatan, Sumatera Barat, Sumatera Utara, Sulawesi Selatan, dan Lampung.
d. Cluster 4 termasuk dalam Provinsi yang memiliki tingkat kasus persebaran COVID-19 yang rendah tingkat kedua. Provinsi yang masuk dalam cluster ini adalah Provinsi Kalimantan Barat, Kalimantan Timur, Kalimantan Tengah, Kalimantan Utara, Kepulauan Riau, Sulawesi

Utara, Sulawesi Tenggara, Sulawesi Tengah, Riau, Maluku Utara, Maluku, Papua Barat, dan Papua.

e. Cluster 5 termasuk dalam Provinsi yang memiliki tingkat kasus persebaran COVID-19 yang rendah tingkat ketiga. Provinsi yang masuk dalam cluster ini adalah Provinsi Aceh, Bangka Belitung, Bengkulu, Jambi, Sulawesi Barat, Nusa Tenggara Timur, dan Gorontalo.

2. Karakteristik cluster yang terbentuk adalah sebagai berikut:

a. Cluster 1 berisi kelompok Provinsi yang memiliki nilai rata-rata kepadatan penduduk, positif, sembuh, dan meninggal di atas rata-rata populasi.

b. Cluster 2 berisi kelompok Provinsi yang memiliki nilai rata-rata kepadatan penduduk, positif, sembuh, dan meninggal di bawah rata-rata populasi.

c. Cluster 3 berisi kelompok Provinsi yang memiliki nilai rata-rata jumlah penduduk, kepadatan penduduk, positif, sembuh, dan meninggal berada di atas rata-rata populasi pada cluster kategori rendah.

d. Cluster 4 berisi kelompok Provinsi yang memiliki nilai rata-rata jumlah penduduk dan positif berada di atas rata-rata populasi pada cluster kategori rendah.

e. Cluster 5 berisi kelompok Provinsi yang memiliki nilai rata-rata positif, sembuh, dan meninggal berada di bawah rata-rata populasi pada cluster kategori rendah.

3. Variabel yang memberikan perbedaan paling besar pada cluster yang terbentuk dalam persebaran kasus COVID-19 adalah variabel kepadatan penduduk dengan nilai F sebesar 26,641 dan nilai signifikan 0,000 dan Provinsi yang memiliki nilai paling besar pada variabel kepadatan penduduk adalah Provinsi DKI Jakarta.

5.2 Saran

Berdasarkan penjelasan dari hasil penelitian di atas, didapatkan beberapa saran, yaitu:

1. Pada penelitian ini lebih banyak ditambahkan informasi berupa grafis supaya pembaca lebih paham mengenai pembahasan dalam penelitian ini.

2. Mengumpulkan data yang lebih lengkap dan terbaru untuk memetakan kelompok kasus COVID-19 dan merancang metode klasifikasi yang lebih akurat.

3. Pada penelitian ini lebih ditambahkan lagi data dan menggunakan metode analisis peramalan untuk memprediksi perkembangan kasus COVID-19 di setiap Provinsi.

4. Untuk mengembangkan penelitian ini, aplikasi yang digunakan dalam analisis *K-Means Cluster*ing bisa menggunakan aplikasi yang lebih modern dan lebih canggih untuk bisa menciptakan visualisasi dari Clustering berupa gambar titik map di setiap Provinsi Indonesia.

5. Untuk pembaca lebih ditingkatkan lagi kewaspadaan dan mengikuti arahan dari Pemerintah. Jika melakukan aktivitas di luar rumah dianjurkan selalu menjaga kebersihan, kesehatan, dan memakai masker terutama di Provinsi yang memiliki jumlah kepadatan penduduk tinggi dan yang berada di klaster tingkat persebaran kasus COVID-19 tinggi.

6. Pemerintah lebih meningkatkan pengawasan, fasilitas, dan penanganan berupa membentuk tim khusus dalam mengobati dan mencegahnya virus COVID-19 dan berperan untuk memberikan harga yang sesuai ke masyarakat dari harga alat, obat, kebutuhan medis, dan pangan supaya ekonomi di masyarakat tidak terbebani oleh pandemik ini

DAFTAR PUSTAKA

Arikunto, S. (2010). *Prosedur Penelitian : Suatu Pendekatan Praktik.* Jakarta: Rineka Cipta .

Bastian, A., Sujadi, H., & Febrianto, G. (2018). *Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)*. *14*(1), 26–32.

BNPB. (2020). Laporan Media Harian Covid19 Tanggal 9 April 2020 Pukul 12.00 Wib. *2020*, *1*, 6575. https://loker.bnpb.go.id/s/GugusTugasCovid19?path=%2FData Kemenkes#pdfviewer

BPS-Statistics Indonesia. (2020). Statistik Indonesia 2020 Statistical Yearbook of Indonesia 2020. *Statistical Yearbook of Indonesia*, *April*. https://doi.org/10.3389/fpsyg.2015.00002

Christiani, C., Tedjo, P., & Martono, B. (2013). *Analisis Dampak Kepadatan Penduduk Terhadap Kualitas Hidup Masyarakat Provinsi Jawa Tengah*. 102–114.

Ediyanto, Novitasari Mara, M., & Satyahadewi, N. (2013). Pengklasifikasian Karakteristik Dengan Metode *K-Means Cluster* Analysis. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*, *02*(2), 133–136.

Ghozali, I. (2018). *Aplikasi Analisis Multivariate Dengan Program IBM SPSS 25 Edisi 9.* Semarang: Badan Penerbit Universitas Diponegoro.

Helilintar, R., & Farida, I. N. (2018). Penerapan Algoritma *K-Means Cluster*ing Untuk Prediksi Prestasi Nilai Akademik Mahasiswa. *Jurnal Sains Dan Informatika*, *4*(2), 80. https://doi.org/10.34128/jsi.v4i2.140

Kemenkes RI. (2020). *Tanya Jawab Seputar Virus Corona*. 119–135.

Mantra, I. B. (2011). *Demografi Umum.* Yogyakarta: Pustaka Belajar.

Pancasa, Sti, R., & Khaerunisa, E. (2018). Analisis Dampak Laju Pertumbuhan Penduduk Terhadap Aspek Kependudukan Berwawasan Gender Pada Urban Area Di Kota Serang. *Tirtayasa Ekonomika*, *13*(1), 130. https://doi.org/10.35448/jte.v13i1.4231

Rauhan, A. (2019). *Pengolahan Data Menggunakan Machine Learning*. *021*, 2–4. https://library.universitaspertamina.ac.id/xmlui/bitstream/handle/123456789/162/Jurnal Ilmiah.pdf?sequence=1&isAllowed=y

Ślusarczyk, B. (2018). Industry 4.0 – Are we ready? *Polish Journal of Management*

Studies, *17*(1), 232–248. https://doi.org/10.17512/pjms.2018.17.1.19

Santoso, S. (2014). *Mahir Statistik Multivariat Dengan SPSS.* Jakarta: PT Elek Media Komputindo.

Simamora, B. (2005). *Analisis Multivariat Pemasaran.* Jakarta: PT Gramedia Pustaka Umum.

Sugiyono. (2011). *Statistika Untuk Penelitian.* Bandung: Alfabeta.

Sukestiyarno. (2016). *Olah Data Penelitian Berbantuan SPSS.* Semarang: UNNES.

Tecuci, G. (2012). Artificial intelligence. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(2), 168–180. https://doi.org/10.1002/wics.200

World Health Organization. (2020). Tatalaksana klinis infeksi saluran pernapasan akut berat (SARI) suspek penyakit COVID-19. *World Health Organization*, *4*(March), 1–25.

Yulianto, S., & Hidayatullah, K. H. (2014). Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat. *Statistika*, *2*(1), 56–63. https://jurnal.unimus.ac.id/index.php/statistik/article/view/1115

LAMPIRAN

Lampiran 3 **Data Tingkat Kasus Persebaran COVID-19 Tanggal 31 Mei 2020 Provinsi Di Indonesia.**

No	Provinsi	X1	X2	X3	X4	X5
1	ACEH	5372	95	20	17	1
2	BALI	4337	750.3	465	329	4
3	BANTEN	12927	1338	861	244	69
4	BANGKA BELITUNG	1489	91	46	27	1
5	BENGKULU	1992	100	91	22	2
6	DI YOGYAKARTA	3843	1206	236	160	8
7	DKI JAKARTA	10558	15900	7348	2082	517
8	JAMBI	3625	72.3	97	15	0
9	JAWA BARAT	49317	1394	2260	619	144
10	JAWA TENGAH	34718	1058.5	1403	361	70
11	JAWA TIMUR	39699	831	4857	609	396
12	KALIMANTAN BARAT	5069	35	189	53	4
13	KALIMANTAN TIMUR	3721	29.8	295	172	3
14	KALIMANTAN TENGAH	2715	18	410	164	19
15	KALIMANTAN SELATAN	4244	109.5	919	99	80
16	KALIMANTAN UTARA	743	10	165	72	2
17	KEPULAUAN RIAU	2190	267	197	102	15
18	NUSA TENGGARA BARAT	5070	252	636	291	10
19	SUMATERA SELATAN	8471	92.5	982	203	34
20	SUMATERA BARAT	5441	129.5	567	265	25
21	SULAWESI UTARA	2507	180	339	51	37
22	SUMATERA UTARA	14563	199.5	409	146	41
23	SULAWESI TENGGARA	2705	71.1	244	122	4
24	SULAWESI SELATAN	8851	193	1541	596	74
25	SULAWESI TENGAH	3054	49	128	61	4
26	LAMPUNG	8448	244	133	71	10
27	RIAU	6972	80.1	117	88	6
28	MALUKU UTARA	1256	39	153	27	11
29	MALUKU	1803	38	223	36	8
30	PAPUA BARAT	960	9.3	168	55	2
31	PAPUA	3379	10.7	675	68	6
32	SULAWESI BARAT	1380	82	92	41	2
33	NUSA TENGGARA TIMUR	5456	114	92	14	1
34	GORONTALO	1203	107	94	26	3

Keterangan:

X1 = Jumlah Penduduk

X2 = Kepadatan Penduduk

X3 = Positif

X4 = Sembuh

X5 = Meninggal

Lampiran 4 **Hasil Ouput SPSS Sebelum Diuji Dengan Analisis *K-Means Cluster***

Uji Multikolinieritas

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Jumlah Penduduk	.434	2.304
	Kepadatan Penduduk	.085	11.813
	Positif	.008	129.797
	Sembuh	.033	30.713
	Meninggal	.013	76.156

a. Dependent Variable: No

Uji Multikolinieritas Dengan Transformasi Data

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	LN_X1	.852	1.174
	LN_X2	.870	1.149
	LN_X3	.347	2.884
	LN_X4	.368	2.721
	LN_X5	.441	2.267

a. Dependent Variable: No

Analisis Deskriptif

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Jumlah Penduduk	34	743	49317	7884.58	11202.153
Kepadatan Penduduk	34	9.3	15900.0	741.059	2708.7090
Positif	34	20	7348	778.00	1470.436
Sembuh	34	14	2082	214.94	371.196
Meninggal	34	0	517	47.44	109.530
Valid N (listwise)	34				

Lampiran 5 **Hasil Ouput SPSS Analisis *K-Means Cluster***

Initial Cluster Centers

Initial Cluster Centers

	Cluster		
	1	2	3
LN_X1	-2.87	-2.30	-5.90
LN_X2	3.44	-11.36	-2.86
LN_X3	2.99	-3.09	-3.95
LN_X4	3.23	-2.36	-6.87
LN_X5	2.91	-1.85	-4.20

Proses Iterasi

Iteration History[a]

	Change in Cluster Centers		
Iteration	1	2	3
1	4.946	5.432	6.270
2	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 2. The minimum distance between initial centers is 10,573.

Final Cluster Centers

Final Cluster Centers

	Cluster		
	1	2	3
LN_X1	-.12	.51	-2.25
LN_X2	.30	-7.48	-2.92
LN_X3	1.50	-.92	-2.41
LN_X4	1.70	-1.37	-2.37
LN_X5	1.33	-.90	-2.37

Perbedaan Variabel pada *Cluster* yang Terbentuk

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
LN_X1	13.671	2	2.815	31	4.857	.015
LN_X2	40.818	2	1.532	31	26.641	.000
LN_X3	16.421	2	2.096	31	7.835	.002
LN_X4	16.222	2	2.157	31	7.522	.002
LN_X5	14.929	2	1.553	31	9.611	.001

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Jumlah Anggota di Setiap *Cluster*

Number of Cases in each Cluster

Cluster	1	2.000
	2	3.000
	3	29.000
Valid		34.000
Missing		.000

Kelompok *Cluster*

Cluster Membership

Case Number	Provinsi	Cluster	Distance
1	ACEH	3	1.841
2	BALI	2	5.432
3	BANTEN	3	4.449
4	BANGKA BELITUNG	3	1.939
5	BENGKULU	3	1.796
6	DI YOGYAKARTA	3	1.673
7	DKI JAKARTA	1	4.946
8	JAMBI	3	1.626
9	JAWA BARAT	1	4.946

10	JAWA TENGAH	2	4.211
11	JAWA TIMUR	2	4.925
12	KALIMANTAN BARAT	3	1.196
13	KALIMANTAN TIMUR	3	2.070
14	KALIMANTAN TENGAH	3	1.840
15	KALIMANTAN SELATAN	3	2.281
16	KALIMANTAN UTARA	3	1.716
17	KEPULAUAN RIAU	3	1.197
18	NUSA TENGGARA BARAT	3	2.517
19	SUMATERA SELATAN	3	6.270
20	SUMATERA BARAT	3	2.475
21	SULAWESI UTARA	3	2.572
22	SUMATERA UTARA	3	3.679
23	SULAWESI TENGGARA	3	1.051
24	SULAWESI SELATAN	3	3.789
25	SULAWESI TENGAH	3	1.269
26	LAMPUNG	3	3.869
27	RIAU	3	2.921
28	MALUKU UTARA	3	1.740
29	MALUKU	3	1.504
30	PAPUA BARAT	3	1.740
31	PAPUA	3	3.030

32	SULAWESI BARAT	3	1.804
33	NUSA TENGGARA TIMUR	3	1.779
34	GORONTALO	3	1.902

Lampiran 6 **Hasil Ouput SPSS Analisis *K-Means Cluster* Pada Cluster 3**

Analisis Deskriptif

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Jumlah Penduduk	29	743	14563	4463.76	3449.558
Kepadatan Penduduk	29	9.3	1338.0	181.459	311.3429
Positif	29	20	1541	348.93	354.817
Sembuh	29	14	596	114.07	121.459
Meninggal	29	0	80	16.62	22.873
Valid N (listwise)	29				

Initial Cluster Centers

Initial Cluster Centers

	Cluster		
	1	2	3
Zscore(LN_X1)	.59580	1.73594	-.00616
Zscore(LN_X2)	.08525	2.28959	-1.73456
Zscore(LN_X3)	-2.43683	1.36454	1.11864
Zscore(LN_X4)	-1.49932	1.23926	-.07419
Zscore(LN_X5)	-1.50243	1.64922	-.16874

Proses Iterasi

Iteration History[a]

	Change in Cluster Centers		
Iteration	1	2	3
1	1.648	1.773	1.603
2	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 2. The minimum distance between initial centers is 4,486.

Final Cluster Centers

Final Cluster Centers

	Cluster		
	1	2	3
Zscore(LN_X1)	-.42593	.98058	-.44952
Zscore(LN_X2)	.07207	.91687	-.67357
Zscore(LN_X3)	-1.20610	.93253	.00384
Zscore(LN_X4)	-1.24873	.99890	-.01916
Zscore(LN_X5)	-1.19416	.98399	-.13008

Perbedaan Variabel pada *Cluster* yang Terbentuk

ANOVA

	Cluster		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Zscore(LN_X1)	6.275	2	.594	26	10.561	.000
Zscore(LN_X2)	6.750	2	.558	26	12.104	.000
Zscore(LN_X3)	9.005	2	.384	26	23.434	.000
Zscore(LN_X4)	9.950	2	.312	26	31.940	.000
Zscore(LN_X5)	8.745	2	.380	25	22.990	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Jumlah Anggota di Setiap *Cluster*

Number of Cases in each Cluster

Cluster	1	7.000
	2	9.000
	3	13.000
Valid		29.000
Missing		.000

Kelompok *Cluster*

Cluster Membership

Case Number	Provinsi	Cluster	Distance
1	ACEH	1	1.648
2	BANTEN	2	1.773
3	BANGKA BELITUNG	1	.844
4	BENGKULU	1	.455
5	DI YOGYAKARTA	2	2.003
6	JAMBI	1	.764
7	KALIMANTAN BARAT	3	1.090
8	KALIMANTAN TIMUR	3	1.249
9	KALIMANTAN TENGAH	3	1.479
10	KALIMANTAN SELATAN	2	1.519
11	KALIMANTAN UTARA	3	2.097
12	KEPULAUAN RIAU	3	1.789
13	NUSA TENGGARA BARAT	2	1.001
14	SUMATERA SELATAN	2	1.055
15	SUMATERA BARAT	2	.761
16	SULAWESI UTARA	3	1.923
17	SUMATERA UTARA	2	1.066
18	SULAWESI TENGGARA	3	.844

19	SULAWESI SELATAN	2	1.739
20	SULAWESI TENGAH	3	.777
21	LAMPUNG	2	1.953
22	RIAU	3	1.665
23	MALUKU UTARA	3	1.427
24	MALUKU	3	.820
25	PAPUA BARAT	3	1.923
26	PAPUA	3	1.603
27	SULAWESI BARAT	1	1.064
28	NUSA TENGGARA TIMUR	1	1.228
29	GORONTALO	1	1.125