

**MENENTUKAN JUMLAH CLUSTER TERBAIK PADA
K-MEAN UNTUK JUMLAH DATA TERJANGKIT
COVID-19**



SKRIPSI

diajukan untuk memenuhi salah satu syarat
memperoleh gelar sarjana komputer (S1)

Oleh :

PUTRI MAULIANA HUSIN

NIM 2016.69.04.0011

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS YUDHARTA PASURUAN

2020

PERNYATAAN PENULIS

JUDUL : MENENTUKAN JUMLAH CLUSTER TERBAIK
PADA K-MEAN UNTUK JUMLAH DATA
TERJANGKIT COVID-19
NAMA : PUTRI MAULIANA HUSIN
NIM : 201669040011

“Saya menyatakan akan bertanggung jawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplikan dan ringkasan yang masing-masing telah jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini adalah karyanya, yang disertai dengan bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut”.

Pasuruan, 16 Agustus 2020



Putri Mauliana Husin
Penulis

PERSETUJUAN SKRIPSI

JUDUL : MENENTUKAN JUMLAH CLUSTER TERBAIK
PADA K-MEAN UNTUK JUMLAH DATA
TERJANGKIT COVID-19
NAMA : PUTRI MAULIANA HUSIN
NIM : 201669040011

Skripsi ini telah diperiksa dan disetujui

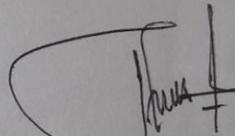
Pasuruan, 16 Agustus 2020

Kaprodi,



M. Imron Rosyadi, M.Kom
NIP. Y. 0690213121

Pembimbing,



Moch. Lutfi, MKom
NIP. Y. 0710048901

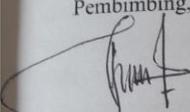
PENGESAHAN SKRIPSI

JUDUL : MENENTUKAN JUMLAH CLUSTER TERBAIK
PADA K-MEAN UNTUK JUMLAH DATA
TERJANGKIT COVID-19
NAMA : PUTRI MAULIANA HUSIN
NIM : 201669040011

Skripsi ini telah diujikan dan dipertahankan di depan Dewan
Penguji pada Sidang Skripsi tanggal 16 Agustus 2020 . Menurut
pandangan kami, Skripsi ini memadai dari segi kualitas untuk
tujuan penganugerahan gelar Sarjana Komputer (S.Kom).

Pasuruan, 16 Agustus 2020

Pembimbing,


Moch. Lutfi, M.Kom

NIP. Y. 0710048901

Penguji Utama,


M. Faishol Amrulloh, M.Kom

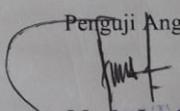
NIP. Y. 0691709007

Kaprodi


M. Imron Rosyidi, M.Kom

NIP. Y. 0690213121

Penguji Anggota,


Moch. Lutfi, M.Kom

NIP. Y. 0710048901

Dekan Fakultas Teknik,


Misbach Munir, ST.MT

NIP. Y. 0690281015

ABSTRACT

At this time the world was shocked by disease new caused by a virus. New virus found at the end of December 2020 in Wuhan China was called the Corona Virus or Covid-19. Within a few months the virus had gone spread throughout the world which causes WHO to set as a world health emergency. In Indonesia alone the case people infected with the Corona virus were only discovered on the 2nd March 2020 as many as 2 people and the number is increasing. In East Java, people have been infected with Covid-19 in the thousands. Of the many infected patients then data clustering is required on the number of people affected by the Corona virus in each region so that the government and the Health Office can identify the distribution in each area. Data Mining is a process used to process large data sets in order to become knowledge with patterns according to the data character. Clustering technique is a record grouping technique in the database based on certain criteria. Therefore this research using the K-Means Clustering method for classifies the spread of the Covid-19 virus in 38 areas in East Java.

Keyword : Covid-19, Clustering, K-Means

ABSTRAK

Pada saat ini dunia sedang digemparkan oleh penyakit baru yang disebabkan oleh virus. Virus baru yang di temukan pada akhir Desember 2020 di Wuhan China disebut Virus Corona atau Covid-19. Dalam beberapa bulan virus tersebut sudah menyebar keseluruh dunia yang menyebabkan WHO menetapkan sebagai kedaruratan Kesehatan dunia. Di Indonesia sendiri kasus orang terjangkit virus Corona baru ditemukan pada tanggal 2 Maret 2020 sebanyak 2 orang dan jumlahnya semakin bertambah. Di Jawa Timur kasus orang terjangkit Covid-19 ini sudah mencapai ribuan. Dari banyaknya pasien terjangkit tersebut maka diperlukan klastering data jumlah orang terjangkit virus Corona pada setiap daerah agar pemerintah dan dinas Kesehatan dapat mengidentifikasi penyebaran di setiap daerahnya. Data Mining merupakan suatu proses yang digunakan untuk mengolah himpunan data besar agar menjadi pengetahuan dengan pola tertentu sesuai dengan karakter data tersebut. Teknik clustering merupakan teknik pengelompokkan record pada basis data berdasarkan kriteria tertentu. Maka dari itu penelitian ini menggunakan metode K-Means Clustering untuk mengelompokkan penyebaran virus Covid-19 di 38 daerah di Jawa Timur.

Kata Kunci : Covid-19, Clustering, K-Means
<G:\PERPUS\KATA PENGANTAR.doc>

Puji syukur kehadiran Allah SWT atas rahmat dan karunia yang telah diberikan kepada penulis, sehingga dapat menyelesaikan skripsi yang berjudul “ Menentukan Jumlah Cluster Terbaik Pada K-Mean Untuk Jumlah Data Terjangkit Covid-19” . Skripsi ini merupakan salah satu syarat untuk memenuhi persyaratan akademis dalam menyelesaikan studi di program Sarjana Teknik Informatika Universitas Yudharta Pasuruan.

Selama mengerjakan skripsi ini, penulis mendapat bantuan dan dukungan dari banyak pihak. Untuk itu, penulis ingin mengucapkan terima kasih kepada:

1. Ibu dan almarhum Ayah yang selalu mendoakan dan tak pernah bosan memberikan semangat dan motivasi kepada putrinya, serta curahan kasih sayang yang tak terhingga sepanjang masa.
2. Romo Kyai Sholeh Bahruddin, selaku pembimbing dan Pembina yang menaungi Universitas Yudharta Pasuruan.
3. Moch. Lutfi, M.Kom selaku dosen pembimbing utama.
4. M. Imron Rosyadi, M.Kom selaku Ketua Program Studi Teknik Informatika Universitas Yudharta Pasuruan.
5. Segenap Bapak dan Ibu dosen yang telah mendidik dan mengajarkan ilmunya kepada penulis selama menempuh pendidikan di Program Studi Teknik Informatika Universitas Yudharta Pasuruan.
6. Segenap staf dan karyawan Program Studi Teknik Informatika Universitas Yudharta Pasuruan.
7. Teman-teman teknik informatika angkatan 2016 yang telah banyak memberikan bantuan, dorongan serta motivasi sehingga skripsi ini dapat terselesaikan.
8. Pihak lain yang telah membantu terselesaikannya skripsi ini yang tidak bisa penulis sebutkan satu-persatu.

Semoga segala pertolongan dan kebaikan semuanya mendapatkan berkah dari Allah SWT. Penulis menyadari bahwa skripsi ini jauh dari kesempurnaan, maka saran dan kritik yang

membangun dari semua pihak sangat diharapkan demi penyempurnaan skripsi ini.

Pasuruan, 16 Agustus 2020

Putri Mauliana Husin

DAFTAR ISI

PERNYATAAN PENULIS	iii
PERSETUJUAN SKRIPSI.....	Error! Bookmark not defined.
PENGESAHAN SKRIPSI	Error! Bookmark not defined.
<i>ABSTRACT</i>	vi
ABSTRAK	vii
KATA PENGANTAR.....	Error! Bookmark not defined.
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
BAB I	1
PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan masalah.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	4
1.6 Sistematika Penulisan.....	4
BAB II	7
TINJAUAN PUSTAKA.....	7
2.1 Penelitian Terkait	7

2.1.1 Analisis Clustering Tingkat Keparahan Penyakit Pasien dengan Menggunakan Algoritma K-Means (Studi Kasus Di Puskesmas Bandar Seikijang) oleh Mentari Tri Indah Rahmayani.....	7
2.1.2 Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk oleh Yulia Darmi dan Agus Setiawan.....	7
2.1.3 Penerapan Algoritma K-Means Clustering Analisis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka) oleh Ade Bastian, Harun Sujadi, dan Gigin Febrianto	7
2.1.4 Optimasi K-Means Clustering Untuk Mengidentifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimization di Kota Semarang oleh Suhardi Rustam, Heru Agus Santoso dan Catur Supriyanto.....	8
2.1.5 Aplikasi Data Mining Untuk Clustering Daerah Penyebaran Penyakit Diare di D Jakarta Menggunakan Algoritma K-Means oleh Teguh Budi Santoso dan Ahmad Sofian Hadi.....	8
2.1.6 Penerapan Algoritma K-Means Clustering Dalam Penentuan Nilai Huruf Pada Permainan Susun Kata Bahasa Bali oleh I Made Rian Yuliawan, Duman Care Khrisne dan Putu Arya Mertasana	8
2.1.7 Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance oleh Muhammad Hariyanto dan Rizky Tahara Shita.....	9

2.1.8 Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan oleh Anindya Khrisna Wardhani	9
2.1.9 Clusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering oleh Parasian D P Silitonga dan Irene Sri Morina.....	9
2.1.10 Clustering Menggunakan Metode K-Means Untuk Penentuan Status Gizi Balita oleh Windha Mega Pradnya Dhuhita.....	10
2.1.11 Coal Trade Data Clusterung Using K-Means (Case Study PT. Global Bangkit Utama) oleh Aulia Tegar Rahman, Wiranto dan Rini Anggrainingsih	10
2.1.12 Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering oleh Ni Putu Eka Merliana, Ernawati dan Alb. Joko Santoso.....	10
2.1.13 Analisis Perbandingan Metode K-Means Dengan Improved SemiSupervised K-Means Pada Data Indeks Pembangunan Manusia (IPM) oleh Gusti Ngurah Wisnu Paramartha, Dian Eka Ratnawat dan Agus Wahyu Widodo	11
2.2 Landasan Teori.....	28
2.2.1 Data Mining.....	28
2.2.2 Clustering	28
2.2.3 Algoritma K-Means Clustering	29
2.2.4 Metode Elbow	31
2.2.5 Metode Silhouette Coefficient.....	32
BAB III.....	36

METODE PENELITIAN	36
3.1 Kerangka Pemikiran	36
3.2 Dataset	37
3.3 Metodologi Penelitian.....	39
3.4 Tahap Pengumpulan Data	39
BAB IV	42
HASIL DAN PEMBAHASAN	42
4.1 Perhitungan Manual K-Means Clustering	42
4.2 Penerapan Pada RStudio	54
BAB V	58
PENUTUP	58
5.1 Kesimpulan.....	58
5.2 Saran.....	58
DAFTAR PUSTAKA.....	60

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	12
Tabel 3.1 Dataset Covid-19.....	37
Tabel 4.1 Data Testing	41
Tabel 4.2 Centeroid 1	42
Tabel 4.3 Hasil Perhitungan Iterasi 1	43
Tabel 4.4 Hasil Pengelompokan Iterasi 1	44
Tabel 4.5 Centeroid 2	44
Tabel 4.6 Hasil Perhitungan Iterasi 2	46
Tabel 4.7 Hasil Pengelompokan Iterasi 2.....	46
Tabel 4.8 Centeroid 3	48
Tabel 4.9 Hasil Perhitungan Iterasi 3	49
Tabel 4.10 Hasil Pengelompokan Iterasi 3.....	49
Tabel 4.11 Centeroid 4	50
Tabel 4.12 Hasil Perhitungan Iterasi 4	51
Tabel 4.13 Hasil Pengelompokan Iterasi 4.....	52

DAFTAR GAMBAR

Gambar 2.1 Flow Chart Algoritma K-Means.....	29
Gambar 3.1 Kerangka Pemikiran	35
Gambar 3.2 Metodologi Penelitian.....	38
Gambar 4.1 Library	53
Gambar 4.2 Input dataset.....	53
Gambar 4.3 Standarisasi Data	53
Gambar 4.4 Menentukan Jumlah Cluster Optimal	53
Gambar 4.5 Metode Wss	54
Gambar 4.6 Metode Silhouette.....	54
Gambar 4.7 Cluster K-Mean	55
Gambar 4.8 Output K-Mean.....	55

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada saat ini dunia digemparkan oleh penyakit baru yang disebabkan oleh virus. Virus tersebut menyebabkan infeksi saluran pernapasan, mulai flu biasa hingga penyakit yang serius seperti *Middle East Respiratory Syndrome* ([MERS](#)) dan Sindrom Pernafasan Akut Berat/ *Severe Acute Respiratory Syndrome* (SARS). Virus jenis baru ditemukan pada manusia sejak kejadian muncul di Wuhan Cina, pada Desember 2019). Coronavirus (coronavirus disease, COVID-19) tanggal 7 Januari 2020 diidentifikasi oleh negara Cina sebagai virus jenis baru. (Pencegahan et al., n.d.)

Gejala umum COVID-19 sama dengan penyakit pernapasan lainnya, termasuk pilek, sakit tenggorokan, batuk, dan demam. Tanpa perawatan khusus 80% penderita COVID-19 dinyatakan pulih. Sekitar 1 dari setiap 6 orang menderita sakit yang parah, seperti disertai pneumonia atau kesulitan bernafas, yang muncul secara bertahap. Pasien yang rentan sakit parah yaitu penderita pada usia lanjut yang memiliki riwayat penyakit seperti diabetes, tekanan darah tinggi dan penyakit jantung. (Pencegahan et al., n.d.)

Jumlah kasus COVID-19 lebih banyak dari pada SARS, tetapi tingkat kematiannya yang lebih rendah yaitu berkisar kurang dari 5% dan SARS di angka 9,6%. Selain itu penyebaran COVID-19 lebih luas ke berbagai negara dibanding SARS. Kasus yang dilaporkan di 192 negara/wilayah sampai dengan tanggal 25 Maret 2020, dikonfirmasi sebanyak 414.179 kasus dengan 18.440 kematian (CFR 4,4%).

WHO telah menetapkan COVID-19 pada tanggal 30 Januari 2020 sebagai Kedaruratan Kesehatan Masyarakat Meresahkan Dunia/ Public Health Emergency of International

Concern (KKMMD/PHEIC). Laporan pertama kasus COVID-19 di Indonesia sebanyak 2 orang yaitu pada tanggal 2 Maret 2020. (Pencegahan et al., n.d.)

Sampai dengan tanggal 10 April 2020, Indonesia sudah melaporkan 3512 kasus konfirmasi COVID-19 dari berbagai provinsi. COVID-19 dapat menular antar manusia melalui percikan batuk dan bersin (droplet). Orang kontak erat dengan pasien COVID-19 dan perawat pasien COVID-19 adalah orang yang sangat berisiko tertular penyakit ini. Cara sederhana untuk mencegah penyebaran infeksi adalah melalui mencuci tangan teratur menggunakan sabun dan air bersih, penerapan etika batuk dan bersin, hindari kontak dengan ternak dan hewan liar serta menghindari siapapun yang menunjukkan gejala penyakit pernapasan. (Pencegahan et al., n.d.)

K-Means merupakan algoritma dalam data mining yang bisa digunakan untuk melakukan pengelompokan/clustering suatu data. Ada berbagai pendekatan pembuatan cluster, diantaranya adalah membuat aturan keanggotaan dalam group berdasarkan tingkat persamaan antara anggota-anggotanya. Pendekatan lainnya adalah membuat sekumpulan fungsi untuk mengukur properti dari pengelompokan tersebut sebagai fungsi dari parameter dari sebuah clustering.

Pada penelitian ini, pengimplementasian menggunakan algoritma clustering K-means. Alasan penggunaan algoritma K-means di antaranya ialah karena algoritma ini memiliki ketelitian cukup tinggi terhadap ukuran objek, sehingga relatif lebih terukur dan efisien dalam pengolahan objek dengan jumlah besar. Algoritma K-means juga tidak terpengaruh urutan objek.

Penyakit epidemic seperti COVID-19 merupakan penyakit yang berbahaya. Maka dari itu pada penelitian ini akan dikaji bagaimana penggunaan algoritma K-Means Cluster pada studi kasus penyakit menular manusia pada suatu objek.

1.2 Rumusan masalah

1. Bagaimana menentukan kelompok tingkat keparahan penyebaran virus Covid-19 pada setiap kabupaten di Jawa Timur ?
2. Bagaimana mendapatkan suatu pengetahuan baru (*knowledge*) terhadap clustering berdasarkan kategori penyebaran virus Covid-19 untuk memudahkan pemerintah dan pelayan Kesehatan dalam meningkatkan kewaspadaan terhadap penyebaran virus penyakit?

1.3 Tujuan Penelitian

1. Mendapatkan Klastering kategori penyakit dengan penyebaran status ODP, PDP dan Positif.
2. Dapat mengelompokkan data daerah penyebaran virus Covid-19 dengan mengimplementasikan algoritma K-Means Clustering maka akan menghasilkan titik-titik pusat penyebaran, dengan itu maka dapat memudahkan pihak Dinas Kesehatan dalam mencegah terjadinya penyebaran virus Covid-19 yang semakin meningkat.
3. Mengetahui persamaan dalam pengolahan data manual menggunakan software.

1.4 Manfaat Penelitian

1. Memberikan informasi pelayan kesehatan masyarakat berdasarkan kategori tingkat keparahan penyebaran virus Covid-19 pada setiap daerah
2. Membantu pemerintah dan dinas Kesehatan agar lebih waspada dan lebih siap dalam menghadapi penyebaran penyakit pandemic

1.5 Batasan Masalah

Untuk mempermudah penelitian ini, dibutuhkan ruang lingkup masalah yang dibahas, diantaranya adalah :

1. Dataset yang digunakan adalah dataset public jumlah penyebaran Covid-19 pada 38 kabupaten di Jawa Timur.
1. Penerapan penelitian ini hanya menggunakan K-Means.
2. Variabel yang digunakan PDP,ODP dan Positif

1.6 Sistematika Penulisan

Agar karya ilmiah ini mudah dipelajari dan dipahami, maka dibagi menjadi lima bab dan setiap bab dibagi menjadi beberapa subbab sesuai topik pembahasan. Sistematika penulisan pada penelitian ini adalah:

Bab 1 Pendahuluan

Membahas mengenai latar belakang penelitian, permasalahan penelitian, tujuan dan manfaat penelitian, dan sistematika penelitian.

Bab 2 Landasan Teori

Berisi tentang tinjauan studi, yaitu membahas tentang penelitian sebelumnya yang mendasari penelitian ini. Dan tinjauan pustaka, yaitu membahas tentang landasan secara teoritis yang diambil dari *textbook*. Serta kerangka pemikiran yang digunakan dalam penelitian ini.

Bab 3 Metodologi Penelitian

Berisi tentang perencanaan penelitian yang terdiri dari analisa kebutuhan, metode pengumpulan data, model yang diusulkan, dan teknik analisis yang dilakukan dalam penelitian Klastering cacat software.

Bab 4 Hasil dan Pembahasan

Berisi hasil penerapan metode K-Means pada dataset penyebaran Covid-19 pada 38 kabupaten di Jawa Timur.

Bab 5 Kesimpulan dan Saran

Berisi kesimpulan dari metode yang diusulkan berdasarkan hasil pengujian, dan saran yang ditujukan untuk pengembangan Klustering menggunakan metode K-Means lebih lanjut.

Halaman ini sengaja dikosongkan

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Penelitian terkait Klastering dengan metode K-Means sudah lama dilakukan, dan sudah banyak yang dipublikasikan dari hasil penelitian. Sebelum dilakukan penelitian, perlu menggali kajian terhadap penelitian sebelumnya, agar dapat mengetahui sejauh mana metode, data, maupun model yang sudah pernah dilakukan. Tinjauan studi ini akan digunakan sebagai landasan penelitian agar dapat mengetahui *state of the art* tentang topik penelitian.

2.1.1 Analisis Clustering Tingkat Keparahan Penyakit Pasien dengan Menggunakan Algoritma K-Means (Studi Kasus Di Puskesmas Bandar Seikijang) oleh Mentari Tri Indah Rahmayani

Dari hasil pengelompokan tingkat keparahan penyakit pasien masyarakat dilingkungan puskesmas Bandar Seikijang lebih banyak menderita penyakit sedang dengan persentase 53,59%. (Rahmayani, 2018)

2.1.2 Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk oleh Yulia Darmi dan Agus Setiawan

Semakin banyak data penjualan barang yang diinput, maka semakin banyak clustering centroid yang terbentuk di proses metode K-means. (Darmi & Setiawan, 2016)

2.1.3 Penerapan Algoritma K-Means Clustering Analisis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka) oleh Ade Bastian, Harun Sujadi, dan Gigin Febrianto

Hasil penelitian yang telah dilakukan dapat disimpulkan bahwa K-means merupakan metode data clustering non hirarki yang mempartisi data ke dalam bentuk satu atau lebih cluster atau kelompok. Kelebihan Algoritma K-Means yaitu sangat cepat dalam pengelompokan objek besar dan pencilan objek. (Bastian et al., n.d.)

2.1.4 Optimasi K-Means Clustering Untuk Mengidentifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimization di Kota Semarang oleh Suhardi Rustam, Heru Agus Santoso dan Catur Supriyanto

Pengujian model daerah sebaran endemik dengan menggunakan k-means dengan optimasi PSO (particle swarm optimization) Lebih baik dari pada K-Means sendiri, bahwa K-Means dengan optimasi PSO (particle swarm optimization) memberikan pemecahan untuk permasalahan identifikasi daerah endemic penyakit menular yang lebih akurat dan bermutu. (Rustam et al., 2018)

2.1.5 Aplikasi Data Mining Untuk Clustering Daerah Penyebaran Penyakit Diare di D Jakarta Menggunakan Algoritma K-Means oleh Teguh Budi Santoso dan Ahmad Sofian Hadi

Pada penelitian ini berhasil menentukan cluster yang teridiri dari C1, C2 dan C3, dimana C1 berjumlah 11 Kelurahan, C2 berjumlah 34 dan C3 berjumlah 43 kelurahan dengan data awal yang digunakan berjumlah 88 data kelurahan. (Santos, 2019)

2.1.6 Penerapan Algoritma K-Means Clustering Dalam Penentuan Nilai Huruf Pada Permainan Susun Kata Bahasa Bali oleh I Made Rian Yuliawan, Duman Care Khrisne dan Putu Arya Mertasana

Penggunaan metode K-Means Clustering dalam penentuan nilai huruf adalah satu cara yang akurat yang dapat digunakan menentukan nilai dari masing-masing huruf dalam permainan susun kata bahasa Bali. (Made et al., 2019)

2.1.7 Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance oleh Muhammad Hariyanto dan Rizky Tahara Shita

Melalui proses pengerjaan dan pengujian dalam penelitian ini, maka dapat ditarik kesimpulan bahwa metode klusterisasi k-means dapat digunakan untuk mengelompokkan daerah potensi. Dimana C1 (Sporadis) berpusat di Kelurahan Rawa Buntu, C2 (Potensi) berpusat di Kelurahan Pondok Ranji dan C3 (Endemis) berpusat di Kelurahan Serua Indah yang ditentukan secara random dari 52 Kelurahan. (Hariyanto & Shita, 2018)

2.1.8 Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan oleh Anindya Khrisna Wardhani

Untuk menentukan konsistensi data kesehatan dapat digunakan teknik data mining yang mampu menggali informasi tersembunyi dari kumpulan data multidimensi yang telah diperoleh, selain itu pengestrakan data yang terhubung dengan data lain juga dapat dilakukan oleh teknik data mining ini. Salah satu teknik data mining yang cukup terkenal yaitu clustering dan metode yang cukup populer dalam teknik data mining ini adalah metode k-means. (Wardhani, 2016)

2.1.9 Clusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering oleh Parasian D P Silitonga dan Irene Sri Morina

Berdasarkan hasil penelitian yang dilakukan disimpulkan bahwa dari sejumlah pasien yang ada, persentasi usia pasien paling tinggi adalah pasien dengan usia tua dan kemudian pasien dengan usia parobaya. (Silitonga & Sri, 2017)

2.1.10 Clustering Menggunakan Metode K-Means Untuk Penentuan Status Gizi Balita oleh Windha Mega Pradnya Duhita

Dengan membandingkan hasil pengelompokan menggunakan tabel Growth Chart dan algoritma K-Means didapat 17 data yang memiliki kelompok yang sama. Dari angka tersebut disimpulkan bahwa algoritma K-Means hanya memiliki nilai akurasi 34% benar. Nilai ini bisa berubah seiring dengan penambahan data latih. (Duhita, 2015)

2.1.11 Coal Trade Data Clustering Using K-Means (Case Study PT. Global Bangkit Utama) oleh Aulia Tegar Rahman, Wiranto dan Rini Anggrainingsih

Kesimpulan yang didapat dari hasil penelitian adalah nilai untuk cluster terbaik dalam proses clustering menggunakan algoritma K-Means yang telah di analisis dan dievaluasi menghasilkan 8 cluster dengan nilai SSE sebesar 0,823066932696445 untuk data total batubara dan nilai SSE sebesar 0,77725237699430 untuk total penjualan batubara. Hasil Clustering yang didapatkan ialah terdapat persamaan hasil cluster pada k-7 yang menunjukkan bahwa cluster tersebut merupakan cluster yang terbaik. (Rahman et al., 2017)

2.1.12 Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering oleh Ni Putu Eka Merliana, Ernawati dan Alb. Joko Santoso

Penentuan jumlah cluster terbaik dengan metode elbow dapat menghasilkan jumlah cluster K yang sama pada jumlah data yang berbeda-beda. (Putu et al., n.d.)

2.1.13 Analisis Perbandingan Metode K-Means Dengan Improved SemiSupervised K-Means Pada Data Indeks Pembangunan Manusia (IPM) oleh Gusti Ngurah Wisnu Paramartha, Dian Eka Ratnawat dan Agus Wahyu Widodo

Pengklusteran data indeks pembangunan manusia menggunakan algoritma semisupervised k-means clustering menggunakan metode silhouette coefficient menghasilkan nilai validasi yang lebih baik dari metode k-means tradisional. Pada lima variasi jumlah data training yaitu 100, 200, 300, 400, dan 500 data, nilai rata-rata yang dihasilkan menggunakan algoritma semi-supervised k-means clustering menghasilkan nilai silhouette coefficient sebesar 0.69880 sedangkan untuk algoritma k-means tradisional menghasilkan silhouette coefficient sebesar 0.62734. (Gusti Ngurah Wisnu Paramartha et al., 2017)

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
1	Mentari Tri Indah Rahmayani	Data penyakit pasien yang berobat pada Puskesmas Bandar Seikijang Kabupaten Pelalawan bulan januari 2017	Analisis Clustering Tingkat Keparahan Penyakit Pasien Dengan Menggunakan Algoritma K-Means (Studi Kasus Di Puskesmas Bandar Seikijang)	K-means Clustering	Dari hasil pengelompokkan tingkat keparahan penyakit pasien masyarakat dilingkungan puskesmas Bandar Seikijang lebih banyak menderita penyakit sedang dengan persentase 53,59%

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
2	Yulia Darmi dan Agus Setiawan	Data produk laku dan produk yang tidak laku, yang dilakukan di Minimarket MM.TI KA Bengkulu dan dilaksanakan pada bulan Juni sampai Juli 2015	Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk	K-means Clustering	Semakin banyak data penjualan barang yang diinput, maka semakin banyak clustering centroid yang terbentuk di proses metode K-means

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
3	Ade Bastian, Harun Sujadi, dan Gigin Febriantoro	Data penyakit menular pada manusia berdasar kan set variabel dibentuk setiap kecamatan tiap Puskesmas yang jumlahnya ada 32 Kantor Puskesmas di Kabupaten Majalengka	Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)	K-means Clustering	Dari hasil penelitian yang telah dilakukan dapat disimpulkan bahwa K-means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok. Kelebihan Algoritma K-means diantaranya adalah mampu mengelompokkan objek besar dan pencilan obyek dengan sangat cepat sehingga mempercepat proses pengelompokan

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
4	Suhardi Rustam, Heru Agus Santoso dan Catur Supriyanto	Data Penderita Penyakit di kota Semarang RS.Karyadi sebagai Rumah Sakit umum dan rujukan di kota Semarang Jawa Tengah	Optimasi K-Means Clustering Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimizati on Di Kota Semarang	K-means Clustering	Pengujian model daerah sebaran endemik dengan menggunakan k-means dengan optimasi PSO (particle swarm optimization) Lebih baik dari pada K-Means sendiri, bahwa K-Means dengan optimasi PSO (particle swarm optimization) memberikan pemecahan untuk permasalahan identifikasi daerah endemic penyakit menular yang lebih akurat dan bermutu

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
5	Teguh Budi Santoso dan Ahmad Sofian Hadi	Data penderita diare dari 88 Kelurahan di DKI Jakarta	Aplikasi Data Mining Untuk Clustering Daerah Penyebaran Penyakit Diare Di Dki Jakarta Menggunakan Algoritma K-Means	K-means Clustering	Pada penelitian ini berhasil menentukan kluster yang terdiri dari C1, C2 dan C3, dimana C1 berjumlah 11 Kelurahan, C2 berjumlah 34 dan C3 berjumlah 43 kelurahan dengan data awal yang digunakan berjumlah 88 data kelurahan

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
6	I Made Rian Yuliawan, Duman Care Khrisne dan Putu Arya Mertasana	Data kata yang digunakan bersumber dari Kamus Bahasa Bali Edisi Ke-2 yang terdiri dari 10441 buah kata	Penerapan Algoritma K-Means Clustering Dalam Penentuan Nilai Huruf Pada Permainan Susun Kata Bahasa Bali	K-means Clustering	Penggunaan metode K-Means Clustering dalam penentuan nilai huruf adalah satu cara yang akurat yang dapat digunakan menentukan nilai dari masing-masing huruf dalam permainan susun kata bahasa Bali

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
7	Muhammad Hariyanto dan Rizky Tahara Shita	Data penduduk dari 52 Kelurahan di kota Tanggerang Selatan pada tahun 2016	Clustering Pada Data Mining Untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean	K-means Clustering	Melalui proses pengerjaan dan pengujian dalam penelitian ini, maka dapat ditarik kesimpulan bahwa metode klasterisasi k-means dapat digunakan untuk mengelompokkan daerah potensi. Dimana C1 (Sporadis) berpusat di Kelurahan Rawa Buntu, C2 (Potensi) berpusat di Kelurahan Pondok Ranji dan C3 (Endemis) berpusat di Kelurahan Serua Indah yang ditentukan secara random dari 52 Kelurahan

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
8	Anindya Khrisna Wardhani	Data penyakit pasien Puskesmas Kajen Pekalongan sebanyak 1000 data	Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan	K-means Clustering	Untuk menentukan konsistensi data kesehatan dapat digunakan teknik data mining yang mampu menggali informasi tersembunyi dari kumpulan data multidimensi yang telah diperoleh, selain itu pengestrakan data yang terhubung dengan data lain juga dapat dilakukan oleh teknik data mining ini. Salah satu teknik data mining terkenal yaitu clustering dan metode yang

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
					cukup populer dalam teknik data mining ini adalah metode k-means

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
9	Parasian D P Silitonga dan Irene Sri Morina	Data rekam medik pasien pengguna BPJS Kesehatan pada bulan Februari tahun 2014	Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering	K-means Clustering	Berdasarkan hasil penelitian yang dilakukan disimpulkan bahwa dari sejumlah pasien yang ada, persentasi usia pasien paling tinggi adalah pasien dengan usia tua dan kemudian pasien dengan usia parobaya

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
10	Windha Mega Pradnya Duhita	Data balita umur dibawah 36 bulan yang ada di POSYAN DU (Pos Pelayanan Terpadu) Desa Karang Songo.	Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita	K-means Clustering	Dengan membandingkan hasil pengelompokan menggunakan tabel Growth Chart dan algoritma K-Means didapat 17 data yang memiliki kelompok yang sama. Dari angka ini disimpulkan bahwa algoritma K-Means hanya memiliki nilai akurasi 34% benar. Nilai ini bisa berubah seiring dengan penambahan data latih

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
11	Aulia Tegar Rahman, Wiranto dan Rini Anggrainingsih	Data penjualan batubara PT Global Bangkit Utama periode Januari 2015 hingga Agustus 2016.	Coal Trade Data Clustering Using K-Means (Case Study PT. Global Bangkit Utama)	K-means Clustering	<p>Nilai untuk cluster terbaik dalam proses clustering menggunakan algoritma K-Means yang telah di analisis dan dievaluasi menghasilkan 8 cluster dengan nilai SSE sebesar 0,823066932696445 untuk data total batubara dan nilai SSE sebesar 0,77725237699430 untuk total penjualan batubara. Hasil clustering terdapat persamaan hasil cluster pada k-7 yang</p>

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
					menunjukkan bahwa cluster tersebut merupakan cluster yang terbaik
12	Ni Putu Eka Merliana , Ernawati dan Alb. Joko Santoso	data selama 3 tahun terakhir untuk mahasiswa yang masih aktif yaitu sebanyak 855 data pada	Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering	K-means Clustering	Penentuan jumlah cluster terbaik dengan metode elbow dapat menghasilkan jumlah cluster K yang sama pada jumlah data yang berbeda-beda

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
		Sekolah Tinggi Agama Hindu Negeri Tampung Penyang Palangka Raya			
13	Gusti Ngurah Wisnu Paramartaha, Dian Eka Ratnawati dan Agus Wahyu Widodo	Data Indeks Pembangunan Manusia (IPM)	Analisis Perbandingan Metode K-Means Dengan Improved SemiSupervised K-Means Pada Data Indeks Pembangunan	K-means Clustering	Pengkusteran data indeks pembangunan manusia menggunakan algoritma semisupervised k-means clustering menggunakan metode silhouette coefficient menghasilkan nilai validasi

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Dataset	Judul	Metode	Hasil
			Manusia (IPM)		yang lebih baik dari metode k-means tradisional. Pada lima variasi jumlah data training yaitu 100, 200, 300, 400, dan 500 data, nilai rata-rata yang dihasilkan menggunakan algoritma semi-supervised k-means clustering menghasilkan nilai silhouette coefficient sebesar 0.69880 sedangkan untuk algoritma k-means tradisional menghasilkan silhouette coefficient sebesar 0.62734

Pada penelitian sebelumnya dengan judul “Implementasi Algoritma K-Means Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Kajen Pekalongan” oleh Anindya Khrisna

Wardhani (Wardhani, 2016) menerapkan metode K-Means pada pengelompokan penyakit pasien dengan menggunakan Rapid Miner sebagai aplikasi pengolah data, sedangkan pada penelitian ini sama-sama pengelompokan data menggunakan K-Means tetapi menggunakan aplikasi RStudio.

2.2 Landasan Teori

2.2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses untuk mengidentifikasi dan mengekstraksi informasi pengetahuan dari beberapa data besar agar lebih bermanfaat dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning. (Silitonga & Sri, 2017)

Data mining hadir dikarenakan kondisi berlimpahnya data (overload data) yang dialami oleh berbagai institusi, perusahaan atau organisasi. Berlimpahnya data ini merupakan akumulasi data transaksi yang terekam bertahun-tahun.(Silitonga & Sri, 2017)

Untuk mempermudah pengolahan data yang besar maka dibutuhkan analisis yang terkomputerisasi menggunakan software data mining yang menunjang arus data dan informasi sesuai dengan kebutuhan dari proses-proses tersebut berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil).(Darmi & Setiawan, 2016)

2.2.2 Clustering

Clustering merupakan suatu teknik yang digunakan untuk menganalisa data baik itu dalam

mesin learning, data mining, pengenalan pola, analisa gambar maupun bioinformatika. Untuk menghasilkan informasi yang bermanfaat diperlukannya suatu analisa data dengan menggunakan proses clustering, ini disebabkan karena data memiliki variasi dan jumlah yang banyak.

Clustering juga bisa dikatakan suatu proses dimana mengelompokkan dan membagi pola data menjadi beberapa jumlah data set sehingga akan membentuk pola yang serupa dan dikelompokkan pada cluster yang sama dan memisahkan diri dengan membentuk pola yang berbeda di cluster yang berbeda. (Putu et al., n.d.)

Peneliti akan menggunakan metode K-Means dimana metode ini menjadi algoritma yang efisien dan efektif dalam mengolah data dalam jumlah yang banyak. Algoritma K-means memiliki permasalahan dalam penentuan jumlah cluster yang terbaik.(Putu et al., n.d.)

2.2.3 Algoritma K-Means Clustering

K-Means merupakan algoritma clustering yang berulang-ulang. Algoritma KMeans dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya cluster yang ingin dibentuk. Nilai-nilai K ditentukan secara random dan nilai tersebut menjadi pusat dari cluster yang disebut dengan centroid, mean atau means.

Metode ini bertujuan untuk meminimalkan jumlah jarak antar semua titik dengan pusat klaster.

Algoritma K-Means diperkenalkan oleh J.B MacQueen pada tahun 1976, yang merupakan salah satu algoritma clustering yang sangat umum untuk mengelompokkan data sesuai dengan karakteristik

atau ciri-ciri bersama yang serupa. Grup data ini dinamakan sebagai cluster.

Adapun langkah-langkah melakukan clustering dengan metode K-Means adalah sebagai berikut (Rahmayani, 2018):

1. Menentukan jumlah cluster k yang akan dibentuk.
2. Inisialisasi k pusat cluster dapat dilakukan dengan berbagai cara. Dapat juga dilakukan penentuan secara random . Pusat-pusat cluster ditentukan nilai awal dengan angka random.
3. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster.
4. r tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat digunakan teori jarak Euclidean yang dirumuskan sebagai berikut

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

dimana:

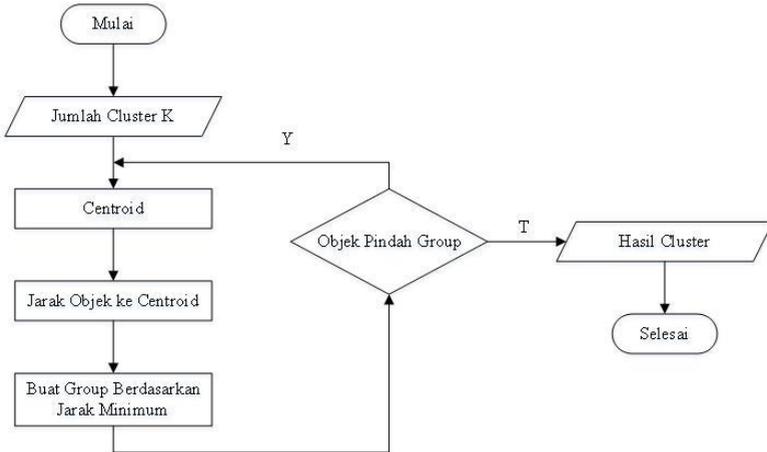
D (i, j) = Jarak data ke i ke pusat cluster j

X ki = Data ke i pada atribut data ke k

X kj = Titik pusat ke j pada atribut ke k

5. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
6. Tugaskan lagi setiap objek memakai pusat cluster yang baru. Jika pusat cluster tidak berubah lagi maka proses clustering selesai.

Kembali ke langkah nomor 3 sampai pusat cluster tidak berubah lagi.



Gambar 2.1 Flow Chart Algoritma K-Means

2.2.4 Metode Elbow

Metode Elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik. (Putu et al., n.d.)

Berikut ini tahapan algoritma metode Elbow dalam menentukan nilai k pada K-Means (Putu et al., n.d.) :

1. Menginisialisasi awal nilai k ;
2. Menaikan nilai k ;
3. Menghitung hasil sum of square error dari tiap nilai k ;
4. Analisis hasil sum of square error dari nilai k yang mengalami penurunan secara drastis ;

5. Cari dan tetapkan nilai k yang berbentuk siku.

Pada metode Elbow nilai cluster terbaik yang akan diambil dari nilai Sum of Square Error (SSE) yang mengalami penurunan yang signifikan dan berbentuk siku. Untuk menghitung SSE menggunakan rumus :

$$SSE = \sum_{K=1}^K \sum_{x_i \in S_k} \| X_i - C_k \|^2$$

Dimana:

K = jumlah cluster

x_i = data ke - i

C_k = centroid cluster

Sum of Square Error (SSE) merupakan rumus yang digunakan untuk mengukur perbedaan antara data yang diperoleh dengan model perkiraan yang telah dilakukan sebelumnya. SSE sering digunakan sebagai acuan penelitian terkait dalam menentukan optimal cluster. (Rahman et al., 2017)

Algoritma Metode Elbow dalam menentukan nilai K pada K-Means (Putu et al., n.d.) :

1. Mulai
2. Inisialisasi awal nilai K
3. Naikkan nilai K
4. Hitung hasil sum of square error dari tiap nilai K
5. Melihat hasil sum of square error dari nilai K yang turun secara drastis
6. Tetapkan nilai K yang berbentuk siku
7. Selesai

2.2.5 Metode Silhouette Coefficient

Metode silhouette coefficient pertama kali dikembangkan oleh (Rousseeuw, 1987) yang mana metode silhouette coefficient merujuk pada validasi dan penafsiran set data. Metode silhouette coefficient

berfungsi untuk mengukur kualitas pada sebuah kluster. Fungsi lain dari silhouette coefficient adalah untuk mengindikasikan derajat kepemilikan setiap objek yang ada di dalam kluster. Metode silhouette coefficient menggabungkan konsep cohesion dan separation sebagai validasi dari hasil klustering.

Untuk menghitung nilai silhouette coefficient diperlukan jarak antar dokumen dengan menggunakan rumus euclidean distance. Setelah itu tahapan untuk menghitung nilai silhouette coefficient adalah sebagai berikut (Gusti Ngurah Wisnu Paramartha et al., 2017):

1. Pada setiap objek data i , hitung nilai rata-rata jarak objek data i dengan seluruh objek data yang berada pada satu kluster yang sama. Nilai rata-rata pada bagian ini dapat didefinisikan dengan $a(i)$. Dalam hal ini rumus dari mencari nilai $a(i)$ adalah:

$$a_{(i)} = \frac{\sum D(i, j)}{|A| - 1}$$

2. Untuk setiap objek data i , hitung nilai rata-rata jarak objek data i dengan semua objek yang ada di kluster lainnya. Dari semua jarak rata-rata yang dihasilkan tersebut, akan diambil rata-rata jarak yang memiliki nilai terkecil. Nilai terkecil ini disebut b_i . Rumus dalam mencari nilai dari b_i adalah sebagai berikut:

$$b_{(i)} = \min(D(i, C))$$

Setelah itu maka untuk objek i memiliki nilai silhouette coefficient:

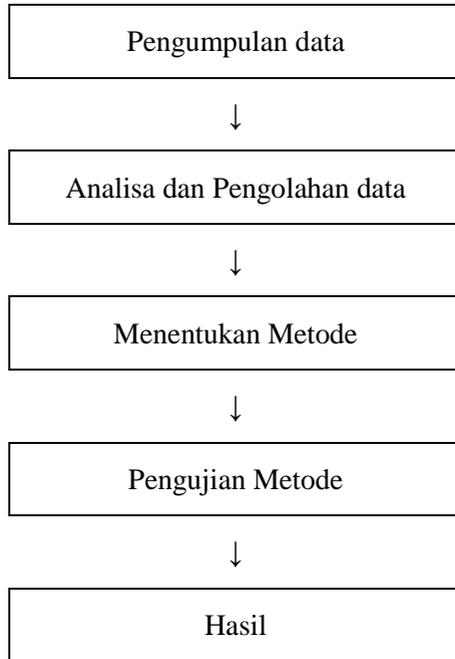
$$S_i = (b_i - a_i) / \max(a_i, b_i)$$

3. Hasil perhitungan nilai silhoutte coeffisien bervariasi dengan rentang -1 sampai 1. Nilai clustering dapat dikatakan baik jika nilai silhoutte coeffisien bernilai positif yaitu ($a_i < b_i$) dan a_i mendekati 0. Dengan hal tersebut menghasilkan nilai silhoutte coeffisien maksimal adalah 1 ketika $a_i = 0$. Jika $s_i = 1$ menandakan bahwa kluster i berada pada kluster yang tepat. Namun jika nilai s_i adalah 0 maka objek i berada antara 2 cluster, maka objek tersebut memiliki struktur tidak jelas. Jika nilai $s_i = -1$ berarti struktur dari kluster memiliki nilai overlapping, dengan itu objek i tepat masuk dalam kluster yang lain. Pada teorinya, nilai rata-rata silhoutte coeffisien dari tiap data objek suatu kluster adalah suatu ukuran yang menunjukkan seberapa ketat data dikelompokkan dalam kluster tersebut. Berikut adalah nilai silhoutte berdasarkan (Gusti Ngurah Wisnu Paramartha et al., 2017):
 1. $0.7 < SC \leq 1$ Struktur yang kuat (strong structure)
 2. $0.5 < SC \leq 0.7$ Struktur yang standar (medium structure)
 3. $0.25 < SC \leq 0.5$ Struktur yang lemah (weak structure)
 4. $SC \leq 0.25$ Tidak memiliki struktur (no structure) (Gusti Ngurah Wisnu Paramartha et al., 2017)

Halaman ini sengaja dikosongkan

BAB III METODE PENELITIAN

3.1 Kerangka Pemikiran



Gambar 3.1. Kerangka Pemikiran

Pada penelitian ini dimulai dari pengumpulan data, data tersebut kita ambil dari dataset public yang sesuai dengan penelitian ini. Setelah data dikumpulkan, kemudian data dibedakan atributnya dan mulai ditentukan metode yang dapat di implementasikan dan di lakukan perhitungan. Setelah itu dilakukan pengujian data dengan memasukkan dataset kedalam system, dan melakukan analisa terhadap system yang telah diuji.

3.2 Dataset

Dataset pada penelitian ini memiliki dua data yaitu data training dan data testing. Data training yaitu data bagian dari dataset yang dilatih untuk menjalankan fungsi dari sebuah algoritma. Sedangkan data testing yaitu bagian dataset yang kita tes untuk melihat keakuratannya, atau dengan kata lain melihat performanya.

Pada skripsi ini menggunakan dataset yang diunduh dari situs sinta.ristekbrin.go.id. Data yang digunakan dalam penelitian ini data jumlah orang terjangkit virus Covid-19 pada 38 daerah di Jawa Timur. Data yang diperoleh terdiri dari 3 atribut yaitu ODP, PDP dan Positif.

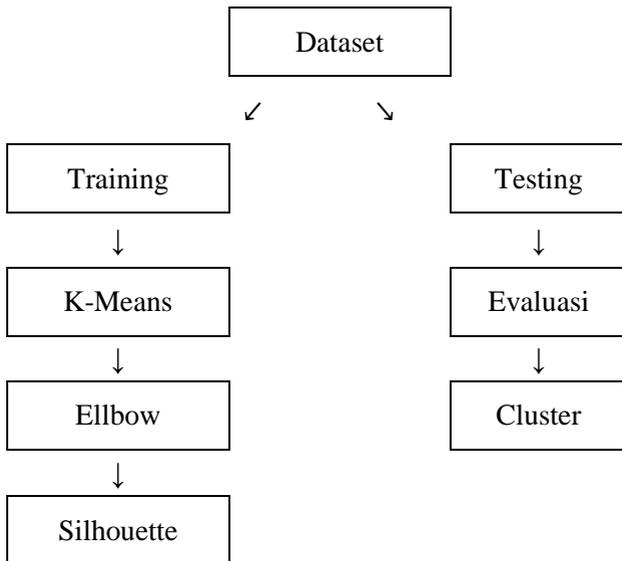
Tabel 3.1 Dataset Covid-19

Kabupaten Kota	ODP	PDP	Positif
Bangkalan	1120	110	348
Banyuwangi	946	79	56
Batu	464	162	162
Blitar	1066	113	150
Blitar City	223	13	36
Bojonegoro	0	11	239
Bondowoso	1407	13	51
Gresik	1684	940	1694
Jember	1355	217	357
Jombang	729	81	507
Kediri	1528	338	391
Kediri City	569	74	107
Lamongan	642	391	310
Lumajang	448	224	128
Madiun	373	130	47
Madiun City	82	33	27
Magetan	435	126	157

Kabupaten Kota	ODP	PDP	Positif
Malang	571	688	475
Malang City	1093	627	573
Mojokerto	587	117	478
Mojokerto – City	868	19	227
Nganjuk	115	223	184
Ngawi	377	75	41
Pacitan	693	28	55
Pamekasan	638	139	229
Pasuruan	1370	453	583
Pasuruan City	155	39	187
Ponorogo	482	84	186
Probolinggo	551	94	199
Probolinggo City	429	72	200
Sampang	549	34	198
Sidoarjo	1626	1044	3080
Situbondo	543	76	216
Sumenep	365	24	211
Surabaya	5121	6737	8489
Trenggalek	968	62	75
Tuban	666	128	199
Tulungagung	1657	739	254

3.3 Metodologi Penelitian

Diagram penelitian dilakukan secara sistematis yang dapat digunakan oleh peneliti sebagai pedoman dalam menyelesaikan penelitian ini agar hasil yang dicapai tidak menyimpang dan tujuan yang diinginkan sesuai dengan yang telah ditetapkan. Adapun Diagram penelitian dalam melakukan penelitian ini adalah sebagai berikut :



Gambar 3.2 Metodologi Penelitian

3.4 Tahap Pengumpulan Data

Jika seorang peneliti berencana melakukan penelitian dengan pendekatan yang lebih kuantitatif, tidak harus memiliki sumber daya dalam hal waktu dan uang, yang tersedia untuk melakukan survei dalam skala besar.

Oleh karena itu, perlu adanya strategi yang lebih bermanfaat yaitu dengan cara analisa data sekunder. Maka pada penelitian ini digunakan data sekunder yang diambil dari situs sinta.ristekbrin.go.id.

Halaman ini sengaja dikosongkan

BAB IV HASIL DAN PEMBAHASAN

4.1 Perhitungan Manual K-Means Clustering

Clustering K-Means merupakan proses dari proses pelatihan. Clustering ini digunakan untuk mengelompokkan data menjadi beberapa cluster sesuai dengan kemiripan yang dimiliki oleh data tersebut. Dalam proses ini data akan dikelompokkan menjadi 3 cluster yaitu cluster ODP, cluster PDP dan cluster Positif.

Langkah-langkah yang dilakukan adalah :

1. Menentukan banyaknya cluster (k)
2. Menentukan centroid
3. Menghitung jarak centeroid
4. Pengelompokan data berdasarkan jarak terdekat
5. Jika objek pindah group maka jumlah cluster bisa ditentukan dan jika tidak maka ulangi proses sampai objek tidak berubah group (stabil)

Iterasi 1

1. Diambil 10 data testing dari 38 data pada tabel 2 dataset covid-19. Dan didapat 3 cluster yaitu OPD, PDP, dan Positif

Tabel 4.1 Data Testing

Kabupaten Kota	ODP	PDP	Positif
Bangkalan	1120	110	348
Banyuwangi	946	79	56
Batu	464	162	162
Blitar	1066	113	150
Blitar City	223	13	36
Bojonegoro	0	11	239
Bondowoso	1407	13	51
Gresik	1684	940	1694

Kabupaten Kota	ODP	PDP	Positif
Jember	1355	217	357
Jombang	729	81	507

2. Dari tabel 3 terpilih secara random 3 data sebagai centeroid / cluster pusat

Tabel 4.2 Centeroid 1

Cluster 1	464	162	162
Cluster 2	223	13	36
Cluster 3	729	81	507

3. Menghitung jarak centeroid

$$d(1,1) = \sqrt{(1120 - 464)^2 + (110 - 62)^2 + (348 - 162)^2} \\ = 683,8392$$

$$d(1,2) = \sqrt{(1120 - 223)^2 + (110 - 13)^2 + (348 - 36)^2} \\ = 954,6528$$

$$d(1,3) = \sqrt{(1120 - 729)^2 + (110 - 81)^2 + (348 - 507)^2} \\ = 423,0875$$

$$d(2,1) = \sqrt{(946 - 464)^2 + (79 - 62)^2 + (348 - 56)^2} \\ = 500,4488$$

$$d(2,2) = \sqrt{(946 - 223)^2 + (79 - 13)^2 + (348 - 56)^2} \\ = 726,2816$$

$$d(2,3) = \sqrt{(946 - 729)^2 + (79 - 81)^2 + (348 - 56)^2} \\ = 500,4938$$

$$d(3,1) = \sqrt{(464 - 464)^2 + (162 - 62)^2 + (162 - 56)^2} \\ = 0$$

$$d(3,2) = \sqrt{(464 - 223)^2 + (162 - 13)^2 + (162 - 56)^2} \\ = 310,0935$$

$$d(3,3) = \sqrt{(464 - 729)^2 + (162 - 81)^2 + (162 - 56)^2} \\ = 442,5054$$

$$d(4,1) = \sqrt{(1066 - 464)^2 + (113 - 62)^2 + (150 - 56)^2} \\ = 604,1101$$

$$d(4,2) = \sqrt{(1066 - 223)^2 + (113 - 13)^2 + (150 - 56)^2} \\ = 856,5308$$

$$d(4,3) = \sqrt{(1066 - 729)^2 + (113 - 81)^2 + (150 - 56)^2} \\ = 491,9766$$

$$d(5,1) = \sqrt{(223 - 464)^2 + (13 - 62)^2 + (36 - 56)^2} \\ = 310,0935$$

$$d(5,2) = \sqrt{(223 - 223)^2 + (13 - 13)^2 + (36 - 56)^2} = 0$$

$$d(5,3) = \sqrt{(223 - 729)^2 + (13 - 81)^2 + (36 - 56)^2} \\ = 694,6229$$

Dari perhitungan tersebut di peroleh hasil sebagai berikut :
Tabel 4.3 Hasil Perhitungan Iterasi 1

No	Kabupaten Kota	Jarak ke Cluster		
		C1	C2	C3
1	Bangkalan	683,8392	954,6528	423,0875
2	Banyuwangi	500,4488	726,2816	500,4938
3	Batu	0	310,0935	442,5054
4	Blitar	604,1101	856,5308	491,9776
5	Blitar City	310,0935	0	694,6229
6	Bojonegoro	493,9899	301,5659	779,8493
7	Bondowoso	961,1301	1184,095	819,9049
8	Gresik	2107,299	2396,417	1748,964
9	Jember	913,7456	1194,186	657,9301
10	Jombang	442,5054	694,6229	0

4. Tempatkan tiap objek ke kluster terdekat berdasarkan nilai centroid yang paling dekat selisihnya (jaraknya)

Tabel 4.4 Hasil Pengelompokan Iterasi 1

No	Kabupaten Kota	Hasil
1	Bangkalan	3
2	Banyuwangi	1
3	Batu	1
4	Blitar	3
5	Blitar City	2
6	Bojonegoro	2
7	Bondowoso	3
8	Gresik	3
9	Jember	3
10	Jombang	3

Iterasi 2

1. Ulangi lagi langkah ke 2 (kedua) seperti Iterasi Ke-1 yaitu untuk mencari perhitungan jarak pusat cluster hingga posisi data tidak mengalami perubahan, dan untuk Iterasi ke-2 jarak pusat cluster nya menjadi seperti berikut,

Tabel 4.5 Centeroid 2

Cluster 1	801,25	108,75	218,75
Cluster 2	111,5	12	137,5
Cluster 3	1391,5	320	612,5

2. Hitung jarak centeroid

$$d(1,1) = \sqrt{(1120 - 705)^2 + (110 - 120.5)^2} \\ + (348 - 109)^2 = 479,0159$$

$$d(1,2) = \sqrt{(1120 - 111.5)^2 + (110 - 12)^2} \\ + (348 - 137.5)^2 = 91034,885$$

$$d(1,3) = \sqrt{(1120 - 1226.833)^2 + (110 - 245.6667)^2} \\ + (348 - 517.8333)^2 = 242,2027$$

$$d(2,1) = \sqrt{(946 - 705)^2 + (79 - 120.5)^2 + (348 - 109)^2} \\ = 250,2244$$

$$d(2,2) = \sqrt{(946 - 111.5)^2 + (79 - 12)^2 + (348 - 137.5)^2} \\ = 841,144$$

$$d(2,3) = \sqrt{(946 - 1226.833)^2 + (79 - 245.6667)^2} \\ + (348 - 517.8333)^2 = 565,6281$$

$$d(3,1) = \sqrt{(464 - 705)^2 + (162 - 120.5)^2 + (162 - 109)^2} \\ = 250,224$$

$$d(3,2) = \sqrt{(464 - 111.5)^2 + (162 - 12)^2 + (162 - 137.5)^2} \\ = 383,8704$$

$$d(3,3) = \sqrt{(464 - 1226.833)^2 + (162 - 245.6667)^2} \\ + (162 - 517.8333)^2 = 845,8913$$

$$d(4,1) = \sqrt{(1066 - 705)^2 + (113 - 120.5)^2} \\ + (150 - 109)^2 = 363,3982$$

$$d(4,2) = \sqrt{(1066 - 111.5)^2 + (113 - 12)^2} \\ + (150 - 137.5)^2 = 959,9102$$

$$d(4,3) = \sqrt{(1066 - 1226.833)^2 + (113 - 245.6667)^2} \\ + (150 - 517.8333)^2 = 422,811$$

$$d(5,1) = \sqrt{(223 - 705)^2 + (13 - 120.5)^2 + (36 - 109)^2} = 499,2086$$

$$d(5,2) = \sqrt{(223 - 111.5)^2 + (13 - 12)^2 + (36 - 137.5)^2} = 150,783$$

$$d(5,3) = \sqrt{(223 - 1226.833)^2 + (13 - 245.6667)^2 + (36 - 517.8333)^2} = 1137,532$$

Dari perhitungan tersebut di peroleh hasil sebagai berikut :
Tabel 4.6 Hasil Perhitungan Iterasi 2

No	Kabupaten Kota	Jarak ke Cluster		
		C1	C2	C3
1	Bangkalan	479,0159	1034,885	242,2027
2	Banyuwangi	250,2244	841,143	565,6281
3	Batu	250,2244	383,8704	845,8913
4	Blitar	363,3982	959,9102	422,811
5	Blitar City	499,2086	150,783	1137,532
6	Bojonegoro	725,2001	150,783	1279,819
7	Bondowoso	712,5477	1298,385	551,8398
8	Gresik	2035,251	2399,298	1440,301
9	Jember	702,3648	1279,257	207,6435
10	Jombang	400,6747	722,9091	524,4716

3. Tempatkan tiap objek ke klaster terdekat berdasarkan nilai centroid yang paling dekat selisihnya (jaraknya)

Tabel 4.7 Hasil Pengelompokan Iterasi 2

No	Kabupaten Kota	Hasil
1	Bangkalan	3

No	Kabupaten Kota	Hasil
2	Banyuwangi	1
3	Batu	1
4	Blitar	1
5	Blitar City	2
6	Bojonegoro	2
7	Bondowoso	3
8	Gresik	3
9	Jember	3
10	Jombang	1

1. Dari hasil iterasi 1 dan iterasi 2 ada perubahan pada posisi data maka dilakukan perhitungan iterasi yang ke 3

Iterasi 3

Tabel 4.8 Centeroid 3

Cluster 1	801,25	108,75	218,75
Cluster 2	111,5	12	137,5
Cluster 3	1391,5	320	612,5

2. Hitung jarak centeroid

$$d(1,1) = \sqrt{(1120 - 801,25)^2 + (110 - 108,75)^2} + (348 - 218,75)^2 = 343,9603$$

$$d(1,2) = \sqrt{(1120 - 111,5)^2 + (110 - 12)^2} + (348 - 137,5)^2 = 1034,885$$

$$d(1,3) = \sqrt{(1120 - 1391,5)^2 + (110 - 320)^2} + (348 - 612,5)^2 = 433,3272$$

$$d(2,1) = \sqrt{(946 - 801,25)^2 + (79 - 108,75)^2} + (348 - 218,75)^2 = 219,8299$$

$$\begin{aligned}
d(2,2) &= \sqrt{(946 - 111,5)^2 + (79 - 12)^2 + (348 - 137,5)^2} \\
&= \mathbf{841,143} \\
d(2,3) &= \sqrt{(946 - 1391,5)^2 + (79 - 320)^2} \\
&\quad + (348 - 612,5)^2 = \mathbf{752,4915} \\
d(3,1) &= \sqrt{(464 - 801,25)^2 + (162 - 108,75)^2} \\
&\quad + (162 - 218,75)^2 = \mathbf{346,1122} \\
d(3,2) &= \sqrt{(464 - 111,5)^2 + (162 - 12)^2 + (162 - 137,5)^2} \\
&= \mathbf{383,8704} \\
d(3,3) &= \sqrt{(464 - 1391,5)^2 + (162 - 320)^2} \\
&\quad + (162 - 612,5)^2 = \mathbf{1043,154} \\
d(4,1) &= \sqrt{(1066 - 801,25)^2 + (113 - 108,75)^2} \\
&\quad + (150 - 218,75)^2 = \mathbf{273,5639} \\
d(4,2) &= \sqrt{(1066 - 111,5)^2 + (113 - 12)^2} \\
&\quad + (150 - 137,5)^2 = \mathbf{959,9102} \\
d(4,3) &= \sqrt{(1066 - 1391,5)^2 + (113 - 320)^2} \\
&\quad + (150 - 612,5)^2 = \mathbf{602,2503} \\
d(5,1) &= \sqrt{(223 - 801,25)^2 + (13 - 108,75)^2} \\
&\quad + (36 - 218,75)^2 = \mathbf{613,9533} \\
d(5,2) &= \sqrt{(223 - 111,5)^2 + (13 - 12)^2 + (36 - 137,5)^2} \\
&= \mathbf{150,783} \\
d(5,3) &= \sqrt{(223 - 1391,5)^2 + (13 - 320)^2 + (36 - 612,5)^2} \\
&= \mathbf{1338,654}
\end{aligned}$$

Dari perhitungan tersebut di peroleh hasil sebagai berikut :
Tabel 4.9 Hasil Perhitungan Iterasi 3

No	Kabupaten Kota	Jarak ke Cluster		
		C1	C2	C3
1	Bangkalan	343,9603	1034,885	433,3272
2	Banyuwangi	219,8299	841,143	752,4915
3	Batu	346,1122	383,8704	1043,154
4	Blitar	273,5639	959,9102	602,2504
5	Blitar City	613,9533	150,783	1338,654
6	Bojonegoro	807,4445	150,783	1473,518
7	Bondowoso	635,7996	1298,385	640,134
8	Gresik	1909,604	2399,298	1280,468
9	Jember	580,9218	1279,257	277,8876
10	Jombang	298,4597	722,9091	712,1499

3. Tempatkan tiap objek ke klaster terdekat berdasarkan nilai centroid yang paling dekat selisihnya (jaraknya)

Tabel 4.10 Hasil Pengelompokan Iterasi 3

No	Kabupaten Kota	Hasil
1	Bangkalan	1
2	Banyuwangi	1
3	Batu	1
4	Blitar	1
5	Blitar City	2
6	Bojonegoro	2
7	Bondowoso	1
8	Gresik	3

No	Kabupaten Kota	Hasil
9	Jember	3
10	Jombang	1

1. Dari hasil iterasi 2 dan iterasi 3 ada perubahan pada posisi data maka dilakukan perhitungan iterasi yang ke 4.

Iterasi 4

Tabel 4.11 Centeroid 4

Cluster 1	955,3333	93	212,3333
Cluster 2	111,5	12	137,5
Cluster 3	1519,5	578,5	1025,5

2. Hitung jarak centeroid

$$d(1,1) = \sqrt{(1120 - 955,33)^2 + (110 - 93)^2 + (348 - 218,75)^2} = 343,9603$$

$$d(1,2) = \sqrt{(1120 - 111,5)^2 + (110 - 12)^2 + (348 - 137,5)^2} = 1034,885$$

$$d(1,3) = \sqrt{(1120 - 1519,5)^2 + (110 - 578,5)^2 + (348 - 1025,5)^2} = 433,3272$$

$$d(2,1) = \sqrt{(946 - 955,33)^2 + (79 - 93)^2 + (348 - 212,33)^2} = 219,8299$$

$$d(2,2) = \sqrt{(946 - 111,5)^2 + (79 - 12)^2 + (348 - 137,5)^2} = 841,143$$

$$d(2,3) = \sqrt{(946 - 1519,5)^2 + (79 - 578,5)^2 + (348 - 1025,5)^2} = 752,4915$$

$$d(3,1) = \sqrt{(464 - 955,33)^2 + (162 - 93)^2} \\ + (162 - 212,33)^2 = 346,1122$$

$$d(3,2) = \sqrt{(464 - 111,5)^2 + (162 - 12)^2 + (162 - 137,5)^2} \\ = 383,8704$$

$$d(3,3) = \sqrt{(464 - 1519,5)^2 + (162 - 578,5)^2} \\ + (162 - 1025,5)^2 = 1043,154$$

$$d(4,1) = \sqrt{(1066 - 955,33)^2 + (113 - 93)^2} \\ + (150 - 212,33)^2 = 273,5639$$

$$d(4,2) = \sqrt{(1066 - 111,5)^2 + (113 - 12)^2} \\ + (150 - 137,5)^2 = 959,9102$$

$$d(4,3) = \sqrt{(1066 - 1519,5)^2 + (113 - 578,5)^2} \\ + (150 - 1025,5)^2 = 602,2503$$

$$d(5,1) = \sqrt{(223 - 955,33)^2 + (13 - 93)^2 + (36 - 212,33)^2} \\ = 613,9533$$

$$d(5,2) = \sqrt{(223 - 111,5)^2 + (13 - 12)^2 + (36 - 137,5)^2} \\ = 150,783$$

$$d(5,3) = \sqrt{(223 - 1519,5)^2 + (13 - 578,5)^2} \\ + (36 - 1025,5)^2 = 1338,654$$

Dari perhitungan tersebut di peroleh hasil sebagai berikut :

Tabel 4.12 Hasil Perhitungan Iterasi 4

No	Kabupaten Kota	Jarak ke Cluster		
		C1	C2	C3
1	Bangkalan	343,9603	1034,8848	433,3272
2	Banyuwangi	219,8299	841,14297	752,4915
3	Batu	346,1122	383,87042	1043,154

No	Kabupaten Kota	Jarak ke Cluster		
		C1	C2	C3
4	Blitar	273,5639	959,91015	602,2504
5	Blitar City	613,9533	150,78296	1338,654
6	Bojonegoro	807,4445	150,78296	1473,518
7	Bondowoso	635,7996	1298,385	640,134
8	Gresik	1909,604	2399,2983	1280,468
9	Jember	580,9218	1279,2566	277,8876
10	Jombang	298,4597	722,90905	712,1499

3. Tempatkan tiap objek ke klaster terdekat berdasarkan nilai centroid yang paling dekat selisihnya (jaraknya)
Tabel 4.13 Hasil Pengelompokan Iterasi 4

No	Kabupaten Kota	Hasil
1	Bangkalan	1
2	Banyuwangi	1
3	Batu	1
4	Blitar	1
5	Blitar City	2
6	Bojonegoro	2
7	Bondowoso	1
8	Gresik	3
9	Jember	3
10	Jombang	1

Karena hasil dari pengelompokan iterasi 3 dan iterasi 4 sudah stabil maka tidak perlu dilakukan iterasi/perulangan lagi.

4.2 Penerapan Pada RStudio

Jumlah data yang akan di analisis ada 38 daerah di Jawa Timur.

Instal package yang akan digunakan,

```
library(factoextra)
library(ggplot2)
library(cluster)
library(NbClust)
library(tidyverse)
```

Gambar 4.1 Library

Input dataset yang digunakan,

```
datacovid=read.delim("clipboard")
```

Gambar 4.2 Input dataset

Lakukan standarisasi data,

```
datafix=scale(datacovid[,2:4])
```

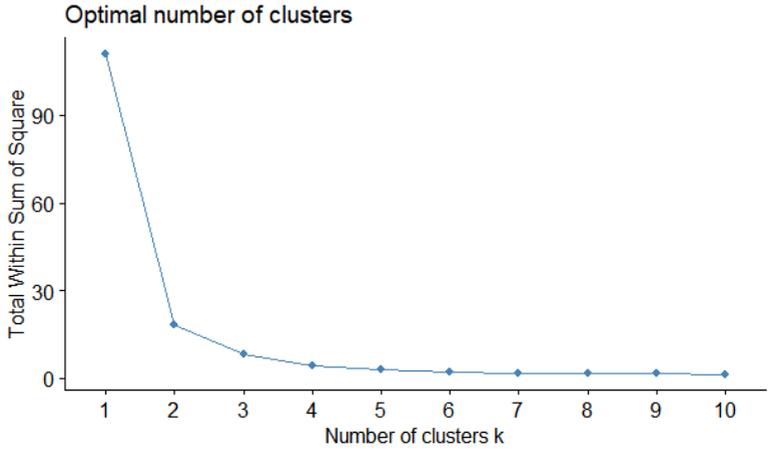
Gambar 4.3 Standarisasi Data

Tentukan jumlah cluster optimal dengan menggunakan syntax sebagai berikut,

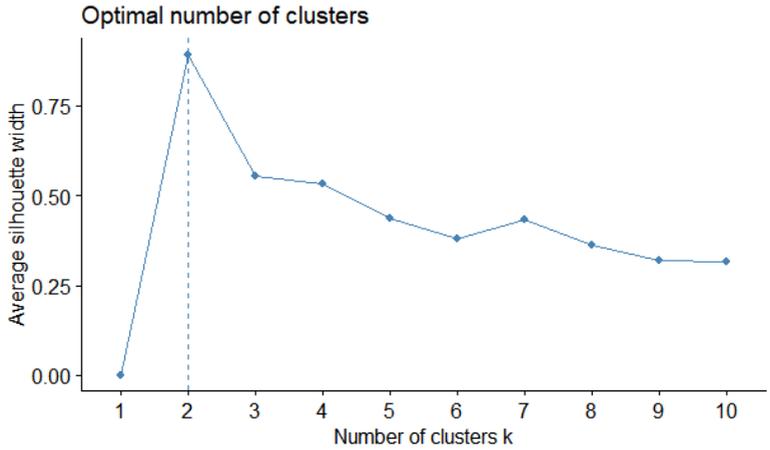
```
fviz_nbclust(datafix, kmeans, method = "wss")
fviz_nbclust(datafix, kmeans, method = "silhouette")
```

Gambar 4.4 Menentukan Jumlah Cluster Optimal

Dan diperoleh output :



Gambar 4.5 Metode Wss



Gambar 4.6 Metode Silhouette

kelompok 3 memiliki 28 anggota. Adapun jarak antar satu cluster yang sama sebesar 92,4 %.

BAB V PENUTUP

5.1 Kesimpulan

Hasil penelitian dari 38 daerah di Jawa Timur dapat disimpulkan bahwa :

1. Pengolahan data dengan menggunakan RStudio diperoleh tiga kelompok data yaitu cluster 1 cluster 1 memiliki 1 anggota, cluster 2 memiliki 9 anggota dan cluster 3 memiliki 38 anggota.
2. Jarak antar satu cluster yang sama sebesar 92,4 %.

5.2 Saran

Berdasarkan pada kesimpulan maka disarankan agar diadakan penelitian lebih lanjut untuk clustering status daerah penyebaran virus Covid-19 di Jawa Timur dengan menambah atribut pada pengolahan data tersebut.

Halaman ini sengaja dikosongkan

DAFTAR PUSTAKA

- Bastian, A., Sujadi, H., & Febrianto, G. (n.d.). *Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). 1*, 26–32.
- Darmi, Y., & Setiawan, A. (2016). Penerapan metode clustering k-means dalam pengelompokan penjualan produk. *Jurnal Media Infotama Universitas Muhammadiyah Bengkulu*, 12(2), 148–157.
- Dhuhita, W. (2015). Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita. *Jurnal Informatika Darmajaya*, 15(2), 160–174.
- Gusti Ngurah Wisnu Paramartha, Dian Eka Ratnawati, & Widodo, A. W. (2017). Analisis Perbandingan Metode K-Means Dengan Improved Semi-Supervised Analisis Perbandingan Metode K-Means Dengan Improved Semi-Supervised K-Means Pada Data Indeks Pembangunan Manusia (IPM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, Vol. 1(January), 813–824. <http://j-ptiik.ub.ac.id>
- Hariyanto, M., & Shita, R. T. (2018). Clustering pada Data Mining untuk Mengetahui Potensi Penyebaran Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode Perhitungan Jarak Euclidean Distance. *Sistem Komputer Dan Teknik Informatika*, 1(1), 117–122.
- Made, I., Yuliawan, R., Care Khrisne, D., & Mertasana, P. A. (2019). Penerapan Algoritma K-Means Clustering dalam Penentuan Nilai Huruf pada Permainan Susun Kata Bahasa Bali. *Jurnal SPEKTRUM*, 6(3), 87–93.
- Pencegahan, P., Pengendalian, D. A. N., Disease, C., & Ke-, R. (n.d.). *Pedoman pencegahan dan pengendalian coronavirus disease (covid-19) revisi ke-4 1*.
- Putu, N., Merliana, E., & Santoso, A. J. (n.d.). *Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means*. 978–979.
- Rahman, A. T., Wiranto, & Rini, A. (2017). Coal Trade Data

- Clustering Using K-Means (Case Study Pt. Global Bangkit Utama). *ITSMART: Jurnal Teknologi Dan Informasi*, 6(1), 24–31. <https://doi.org/10.20961/ITS.V6I1.11296>
- Rahmayani, M. T. I. (2018). Analisis Clustering Tingkat Keperahan Penyakit Pasien Menggunakan Algoritma K-Means. *Jurnal Inovasi Teknik Informatika*, 1(2), 40–44.
- Rustam, S., Santoso, H. A., & Supriyanto, C. (2018). Optimasi K-Means Clustering Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimization Di Kota Semarang. *ILKOM Jurnal Ilmiah*, 10(3), 251–259. <https://doi.org/10.33096/ilkom.v10i3.342.251-259>
- Santos, T. B. (2019). Aplikasi Data Mining untuk Clustering Daerah Penyebaran Penyakit Diare di DKI Jakarta Menggunakan Algoritma K-MEANS. *Jurnal Ilmiah FIFO*, 11(2), 131. <https://doi.org/10.22441/fifo.2019.v11i2.003>
- Silitonga, P. D. P., & Sri, I. (2017). Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering. *TIMES (Technology Informatics & Computer System)*, VI(2), 22–25.
- Wardhani, A. K. (2016). Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan. *Jurnal Transformatika*, 14(1), 30–37.