

**ANALISIS CLUSTER PADA KELOMPOK MASYARAKAT YANG
RENTAN TERHADAP PAPARAN COVID-19 MENGGUNAKAN
METODE K-MEANS CLUSTERING DAN VISUALISASI DENGAN SIG**

SKRIPSI

Diajukan untuk Memenuhi Salah Satu Syarat Kelulusan

Jenjang Strata 1 pada Program Studi Teknik Informatika

Universitas Jenderal Achmad Yani



Oleh :

Indra Raja DrI

34111711139

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN INFORMATIKA

UNIVERSITAS JENDERAL ACHMAD YANI

2022

LEMBAR PENGESAHAN

ANALISIS CLUSTER PADA MASYARAKAT YANG RENTAN TERHADAP PAPARAN COVID-19 MENGGUNAKAN METODE K- MEANS CLUSTERING DAN VISUALISASI DENGAN SIG

“Setelah membaca skripsi ini dengan seksama, menurut pertimbangan kami telah memenuhi persyaratan ilmiah sebagai suatu skripsi”

Cimahi, Januari 2021

Indra Raja Drl

34111711139

Menyetujui,

Pembimbing 1

Pembimbing 2



**Yulison Herry Chrisnanto,
S.T., M.T.**

NID. 4121 668 63

**Fajri Rakhmat U, S.T., M.T.
NID. 4121 858 88**

Mengetahui,

Dekan Fakultas
Sains dan Informatika

Ketua Jurusan Informatika

**Dr. Anceu Murniati, S.Si.,
M.Si.
NID. 412126369**

**Agus Komarudin S.kom.,
M.T.
NID. 4121 758 78**

PERNYATAAN KEASLIAN

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **“ANALISIS CLUSTER PADA KELOMPOK MASYARAKAT YANG RENTAN TERHADAP PAPARAN COVID-19 MENGGUNAKAN METODE K-MEAN CLUSTERING DAN VISUALISASI DENGAN SIG”** ini beserta isinya adalah benar-benar karya saya, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung resiko atau sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya atau ada klaim dari pihak lain terhadap keaslian karya saya ini.

Cimahi, .. 2021

Pembuat Pernyataan



Indra Raja Dr1

3411171139

ABSTRAK

Penyakit baru Coronavirus-2019 telah menimbulkan bahaya. Ini adalah penyakit yang menular dan kecepatan serta laju infeksi yang cepat penularan di antara orang-orang. Pada 26 Maret 2020, Amerika Serikat menjadi negara terdepan dijumlah kasus di seluruh dunia, menggantikan Italia yang sebelumnya masuk pemimpin kasus COVID-19. Pada 12 April 2020, 20.000 kematian dan 500.000 kasus telah dikonfirmasi. Selain beban penyakit langsung dari Covid-19, respon pandemi sudah menyebabkan efek secara tidak langsung negatif. Ini ditanggung secara tidak proporsional oleh orang-orang yang sudah memiliki Biaya keuangan lebih sedikit dan kesehatan yang lebih buruk, orang yang sebelumnya sudah memiliki gangguan kesehatan dan penggunaan zat akan meningkatkan risiko terinfeksi Covid-19, pada penelitian ini menggunakan metode k-mean clustering dan Geographic Information System untuk memetakan sebaran jenis penyakit berdasarkan data yang membawa terhadap kerentanan suatu wiliyah terpaparnya virus. Untuk dapat dilakukan analisis kesehatan apakah suatu wilayah rentan terhadap terpaparnya covid-19 sehingga upaya kesehatan dapat dilakukan.

Kata kunci : *k-means clustering, Covid-19, Sistem informasi Geografis*

ABSTRACT

The new disease Coronavirus-2019 has posed a danger. It is a contagious disease and the speed and fast rate of infection is transmission between people. On March 26, 2020, the United States became the leading country in the number of cases worldwide, replacing Italy, which was previously the leader of COVID-19 cases. As of April 12, 2020, 20,000 deaths and 500,000 cases have been confirmed. In addition to the direct disease burden of Covid-19, the pandemic response has already caused negative indirect effects. This is disproportionately borne by people who already have less financial costs and poorer health, people who already have health problems and substance use will increase the risk of being infected with Covid-19, in this study using the k-mean clustering method and Geographic Information System to map the distribution of disease types based on data that leads to the vulnerability of a region to being exposed to the virus. To be able to do a health analysis of whether an area is susceptible to exposure to COVID-19 so that health efforts can be carried out.

Keywords: *k-means clustering, Covid-19, Geographic Information System*

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT, yang telah memberi rahmat, hidayah dan karunia-Nya kepada penulis dalam usaha untuk menyelesaikan penelitian yang berjudul "**ANALISIS CLUSTER PADA KELOMPOK MASYARAKAT YANG RENTAN TERHADAP PAPARAN COVID-19 MENGGUNAKAN METODE K-MEAN CLUSTERING DAN VISUALISASI DENGAN SIG**". Shalawat dan salam selalu tercurah limpahkan untuk junjunan kita Nabi Muhammad SAW, keluarga, para sahabat dan para penegak risalah hingga akhir zaman.

Penelitian ini disusun sebagai salah satu syarat untuk memenuhi kelulusan jenjang Strata 1 pada program studi informatika Universitas Jenderal Achmad Yani.

Akhir kata, semoga penelitian ini dapat bermanfaat khususnya bagi penulis, serta seluruh pihak yang berkepentingan pada umumnya. Semoga Allah SWT selalu melimpahkan karunia-Nya pada kita semua. Aamiin.

Cimahi, .. 2021

Pembuat Pernyataan



Indra Raja Drl

3411171139

UCAPAN TERIMAKASIH

Selama proses penyusunan skripsi ini, terdapat banyak permasalahan serta hambatan yang dihadapi penulis. Namun, dengan adanya kekuatan dan kemudahan yang diberikan oleh Allah SWT, selain itu terdapat pula masukan, bantuan dan beberapa dukungan baik secara moril maupun material yang diberikan dari berbagai pihak sebagai faktor penting demi menyelesaikan penelitian ini. Maka dari itu pada kesempatan ini penulis dengan segala kerendahan hati dan ketulusan hati ingin menyampaikan penghargaan dan ucapan terimakasih sebesar-besarnya kepada :

1. Bapak Rizal Septamara dan Ibu Siti Dinar selaku orang tua serta Indriani putri selaku kaka dan adik indira rizki tidak lupa keluarga besar yang selalu memberikan kasih sayang, do'a dan dukungan penuh kepada penulis.
2. Yth. Ade Kanianingsih, S.Si., M.Stat., selaku Dosen Wali penulis yang selalu memberikan motivasi serta dukungan terhadap penulis.
3. Yth. Bapak Yulison H. Chrisnanto, S.T., M.T. selaku Dosen Pembimbing 1 dan Fajri Rakhmat Umbara, S.T., M.T selaku Dosen Pembimbing 2 yang telah memberikan masukan serta pengetahuannya serta selalu sabar dalam membimbing dan mengarahkan penulis selama penyusunan penelitian ini.
4. Yth. Dr. Anceu Murniati, S.Si., M.Si. selaku Dekan Fakultas Sains dan Informatika Universitas Jenderal Achmad Yani.
5. Yth. Agus Komarudin, S.Kom, M.T. selaku Ketua Jurusan Informatika Universitas Jenderal Achmad Yani.
6. Seluruh dosen beserta staff karyawan di Jurusan Informatika yang telah mendidik dan membekali ilmu bagi penulis dengan ilmu pengetahuan yang sangat berharga selama masa perkuliahan.

Akhir kata semoga amal baik Bapak/Ibu/Kerabat sekalian mendapat balasan dan karunia Allah SWT lebih dari apa yang diterima penulis.

Cimahi januari,2022

Pembuat pernyataan



Indra Raja Drl

3411171139

DAFTAR ISI

LEMBAR PENGESAHAN.....	I
PERNYATAAN KEASLIAN.....	II
ABSTRAK.....	III
<i>ABSTRACT.....</i>	IV
KATA PENGANTAR.....	V
UCAPAN TERIMAKASIH.....	VI
DAFTAR ISI.....	VIII
DAFTAR GAMBAR.....	XI
DAFTAR TABEL.....	XII
DAFTAR SIMBOL.....	XIII
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah.....	3
1.4. Tujuan penelitian.....	4
1.5. Luaran dan Manfaat Penelitian.....	4
1.6. Metode Penelitian.....	4
1.7. Sistematik Penelitian.....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1. Studi Literatur.....	8
2.1.1. Pengelompokan provinsi terdampak pandemi covid-19 di Indonesia.....	8
2.1.2. Correlation Between Temperature and COVID-19 (Suspected, Confirmed and Death).....	8
2.1.3. Identification of spatial variations in COVID-19 epidemiological data using K-Means clustering algorithm.....	9
2.2. Landasan Teori.....	9
2.2.1. K-Mean Clustering.....	9
2.2.2. Komparasi Metode Clustering.....	11
2.2.3. Elbow Method.....	12
2.2.4. Silhouette Coefficient.....	14

2.2.5. Min-Max Scaler.....	15
2.2.6. Sistem Informasi Geografis.....	16
2.2.7. Rata-Rata Variable.....	17
2.2.8. Unified Model Language.....	17
2.3. Variable Penelitian.....	20
2.3.1. Obesitas.....	20
2.3.2. Diabetes.....	20
2.3.3. Tuberculosis.....	20
2.3.4. Hipertensi.....	21
2.3.5. Diare.....	21
2.3.6. Jumlah Penduduk.....	21
2.3.7. Covid-19.....	22
BAB III ANALISIS DAN PERANCANGAN.....	23
3.1. Analisis Sistem Berjalan.....	23
3.2. Perancangan Sistem Analisis kesehatan masyarakat yang rentan terhadap paparan covid-19 menggunakan metode k-mean clustering dan visualisasi sig.....	24
3.3. Pre-processing Data komorbid covid-19.....	25
3.3.1. Remove Duplicate.....	26
3.3.2. Data Integration.....	30
3.3.3. Min-Max Scaler.....	33
3.3.4. Feature Selection.....	36
3.4. Metode Elbow.....	37
3.5. K-mean Clustering.....	42
3.6. Rata-Rata Variable.....	49
3.7. Perancangan Perangkat Lunak.....	50
3.7.1. Perancangan Antarmuka Dashboard.....	71
3.7.2. Perancangan Antarmuka Import File.....	72
3.7.3. Perancangan Antarmuka Elbow.....	73
3.7.4. Perancangan Antarmuka Input jumlah cluster.....	74
3.7.5. Perancangan Antarmuka Proses K-means.....	75

3.7.6. Perancangan Antarmuka Hasil Clustering.....	76
BAB IV IMPLEMENTASI DAN PENGUJIAN.....	78
4.1. Implementasi Sistem.....	78
4.1.1. Implementasi Antarmuka dashboard.....	78
4.1.2. Implementasi Antarmuka import file.....	79
4.1.3. Implementasi Antarmuka elbow.....	79
4.1.4. Implementasi Antarmuka input cluster.....	81
4.1.5. Implementasi Antarmuka Proses K-means.....	82
4.1.6. Implementasi Antarmuka Hasil clustering.....	84
4.2. Pengujian Perangkat Lunak.....	85
4.3. Pengujian Cluster.....	93
BAB VKESIMPULAN DAN SARAN.....	95
5.1. Kesimpulan.....	95
5.2. Saran.....	95
REFERENSI.....	97

DAFTAR GAMBAR

1 .1 Gambar Knowledge Discovery Data.....	5
2.1 Gambar K-mean Clustering.....	11
2.2 Gambar Elbow method.....	13
2.3 Gambar Clustering Map.....	16
2 4 Gambar Sum of Square Error.....	38
3 1 Gambar Use Case Diagram.....	53
3 2 Gambar Activity Diagram preprocessing.....	63
3 3 Activity Diagram Elbow method.....	63
3 4 Activity Diagram K-mean Clustering.....	65
3 5 Activiy Diagram Visualiasi MAP.....	66
3 6 Sequence Diagram Preprocessing.....	67
3 7 Sequence Diagram Elbow.....	68
3 8 Sequence Diagram K-mean Clustering.....	69
3 9 Sequence Diagram Visualiasi Map.....	70
3 10 Class Diagram Sistem.....	71
3.11 Gambar Perancangan Antarmuka Dashboard.....	72
3.12 Gambar Perancangan Antarmuka Import File.....	73
3.13 Gambar Perancangan Antarmuka Elbow.....	74
3.14 Gambar Perancangan Antarmuka Input Jumlah cluster.....	75
3.15 Gambar Perancangan Antarmuka Proses K-Means.....	76
3.16 Gambar Perancangan Antarmuka Hasil Clustering.....	77

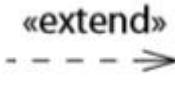
4.1 Gambar Implementasi Antarmuka.....	78
4.2 Gambar Implementasi Antarmuka Import File.....	79
4.3 Gambar Implementasi Antarmuka Elbow 1.....	80
4.4 Gambar Implementasi Antarmuka Elbow 1.....	81
4.5 Gambar Implementasi Antarmuka Input Cluster.....	82
4.6 Gambar Implementasi Antarmuka Perulangan-I 1.....	83
4.7 Gambar Implementasi Antarmuka Perulangan -I 1.....	84
4.8 Gambar Implementasi Antarmuka Hasil clustering.....	85

DAFTAR TABEL

3. 1 Table Dengan Variable Duplikasi.....	26
3. 2 Table Setelah Dikakukan Elimnasi Variable.....	28
3. 3 Table Dengan Variable Terpisah.....	30
3. 4 Table yang telah dilakukan Data Integrasi.....	32
3. 5 Table Min-Max Scaler interval 0-2.....	33
3. 6 Table Feature Selection Variable yang tidak digunakan.....	36
3. 7 Table Fungsional Perangkat Lunak.....	51
3. 8 Table Use Case Scenario Remove Duplicate.....	54
3. 9 Table Use Case Scenario data integration.....	55
3. 10 Table Use Case Scenario Min-max scaler.....	56
3. 11 Table Use Case Scenario Feature Selection.....	57
3. 12 Table Use Case Scenario elbow method.....	58
3. 13 Table Use Case Scenario k-mean.....	60
3. 14 Table Use Case Scenario visualiasi map.....	61
 4.1 Tabel Tujuan Pengujian.....	86
4.2 Table Pengujian Perangkat Lunak.....	88
4.3 Table Pengujian Fungsional Perangkat Lunak.....	90

DAFTAR SIMBOL

1. Simbol Use Case Diagram

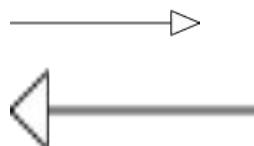
No	Simbol	Keterangan
1.	 Aktor	Aktor merupakan pengguna atau sistem lain yang berinteraksi dengan sistem yang dibuat di luar sistem yang dibuat itu sendiri, jadi walaupun simbol aktor adalah gambar orang, tetapi aktor belum tentu merupakan orang, dinyatakan dengan kata benda di awal frase.
2.		Use Case fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit atau aktor, biasanya dinyatakan dengan menggunakan kata kerja di awal frase nama Use Case.
3.		Association menunjukkan hubungan antara aktor dengan Use Case atau antar Use Case dengan Use Case.
4.		Dependency adalah sebuah relasi yang menunjukkan antara dua model elemen.
5.		Extends adalah relasi Use Case tambahan ke sebuah Use Case dan Use Case yang ditambahkan dapat berdiri sendiri walaupun tanpa sebuah Use Case tambahan relasi tersebut.
6.		Include adalah relasi Use Case tambahan ke sebuah Use Case dan Use Case yang ditambahkan memerlukan Use Case tersebut untuk menjalankan fungsinya atau sebagai syarat dijalankannya Use Case tersebut.
7.		Generalization merupakan hubungan generalisasi dan spesialisasi (umum-khusus) antara dua buah Use Case dan fungsi yang satu adalah fungsi yang lebih umum.

2. Simbol Activity Diagram

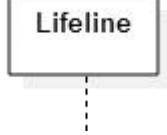
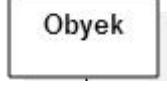
No	Simbol	Keterangan
1.		Initial node merupakan status awal pada aktivitas sistem.
2.		Activity merupakan aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja.
3.		Status akhir merupakan aktivitas terakhir yang dilakukan sistem.
4.		Join atau penggabungan merupakan asosiasi lebih dari satu aktivitas yang digabungkan menjadi satu.
5.		Decision merupakan asosiasi percabangan digunakan jika ada pilihan aktivitas lebih dari satu.
6.		State atau status adalah keadaan sistem pada waktu tertentu. State dapat berubah jika ada kejadian tertentu yang memicu perubahan tersebut.
7.	Event	Event adalah kegiatan yang menyebabkan berubahnya status mesin.
8.		Choice Point menunjukkan kondisi atau keputusan memilih salah satu state.

3. Simbol Class Diagram

No	Simbol	Keterangan
1.		Kelas pada struktur sistem yang terdiri dari nama kelas, atribut kelas, dan operasi yang terdapat pada kelas tersebut.
2.		Asosiasi merupakan relasi antar kelas dengan makna umum, asosiasi biasanya juga disertai dengan multiplicity.

No	Simbol	Keterangan
3.		Generalisasi merupakan relasi antar kelas dengan makna generalisasi-spesialisasi (umum - khusus).
4.		Hubungan asosiasi satu arah dari satu kelas ke kelas lain dalam sistem.
5.		Agregasi menunjukkan suatu kelas adalah bagian dari kelas yang lain.
6.		Komposisi bentuk khusus dari agregasi dimana kelas yang lain menjadi bagian diciptakan.
7.		Hubungan ketergantungan satu arah dari satu kelas ke kelas lain dalam sistem.
8.		Relasi hubungan suatu kelas merupakan realisasi dari kelas lain.

4. Simbol Sequence Diagram

No	Simbol	Keterangan
1.		Aktor merupakan orang, proses atau sistem lain yang berinteraksi dengan sistem yang dibuat.
2.		Lifeline menyatakan kehidupan suatu obyek.
3.		Obyek menyatakan obyek yang berinteraksi dengan pesan.
4.		Waktu aktif, menyatakan obyek dalam keadaan aktif dan berinteraksi pesan.
5.	1: nama_metode()	Pesan tipe call, menyatakan suatu obyek yang memanggil operasi/ metode yang ada pada obyek lain atau dirinya sendiri. Arah panah mengarah pada obyek yang

No	Simbol	Keterangan
		memiliki metode.
6.	1: masukan 	Pesan tipe send, menyatakan bahwa suatu obyek mengirimkan data atau informasi ke obyek lainnya, arah panah mengarah pada obyek yang dikirim.
7.	1: keluaran 	Pesan tipe return, menyatakan bahwa suatu obyek yang telah menjalankan suatu operasi atau metode menghasilkan suatu kembalian ke obyek tertentu, arah panah mengarah pada obyek yang menerima kembalian.
8.	Pesan tipe destroy 	Menyatakan suatu obyek mengakhiri hidup obyek yang lain, arah panah mengarah pada obyek yang diakhiri sebaiknya jika ada create maka ada destroy.

BAB I PENDAHULUAN

1.1. Latar Belakang

Penyakit baru Coronavirus-2019 telah menimbulkan bahaya yang luar biasa bagi seluruh umat manusia di seluruh dunia. Ini adalah penyakit manusia yang menular dan kecepatan serta laju infeksi yang cepat penularan di antara orang-orang telah memperkenalkan tantangan baru bagi sistem perawatan kesehatan[1]Tanggal 30 Januari 2020, WHO menyatakan wabah penyakit di Cina COVID-19 menjadi Darurat Kesehatan Masyarakat yang Menjadi Perhatian Internasional menimbulkan risiko tinggi bagi negara-negara dengan sistem kesehatan yang rentan[5].COVID-19 menyebar kenegara yang berbeda pun bervariasi tergantung pada berbagai faktor yaitu; faktor iklim seperti suhu atau migrasi orang[4]. Pada tahap awal, jumlah terinfeksi kasus sedikit meningkat, namun, setelah mencapai 100 kasus yang dikonfirmasi terinfeksi, tingkat pertumbuhannya cepat atau sedikit ada perubahan[4].Dengan cepatnya penyebaran wabah COVID-19, orang sangat prihatin dengan penyebarannya, keparahan, dan efek negatif yang luar biasa pada kesehatan publik dan masyarakat[3].Pada 26 Maret 2020, Amerika Serikat menjadi negara terdepan dijumlah kasus di seluruh dunia, menggantikan Italia yang sebelumnya masuk pemimpin kasus COVID-19[2]. Pada 12 April 2020, 20.000 kematian dan 500.000 kasus telah dikonfirmasi di Amerika Serikat [2.]di Thailand lebih dari 10 jutaan orang akan menghadapi pengangguran[4].

Selain beban penyakit langsung dari Covid-19, respon pandemi sudah menyebabkan efek secara tidak langsung negatif. Ini ditanggung secara tidak proporsional oleh orang-orang yang sudah memiliki Biaya keuangan lebih sedikit dan kesehatan yang lebih buruk[7].orang yang sebelumnya sudah memiliki gangguan kesehatan dan penggunaan zat akan meningkatkan risiko terinfeksi Covid-19, [6].gangguan kesehatan ini dikatikan dengan Penyakit

tuberkulosis dengan peningkatan 2,10 kali lipat risiko penyakit COVID-19 yang parah[10].tuberkulosis memiliki peluang lebih tinggi untuk mengembangkan komplikasi serius dari COVID-19[11]Diabetes melitus dikaitkan perkembangan penyakit pada pasien dengan COVID-19[12]adapun masyarakat yang di diagnosis memiliki obsitas menjadi pengaruh terinfeksi covid-19 dan membawanya lebih parah[17].Saat pasien memiliki diare,merupakan gejala yang dapat terkonfirmasi karena pandemi covid-19[18].

Ada pun Faktor lain yang mempengaruhi penyebaran covid-19 seperti[19] kepadatan penduduk yang tinggi di kota besar dan daerah metropolitan merupakan faktor risiko utama untuk tertular virus.

Analisis kesehatan masyarakat perlu dilakukan untuk mengintrupsi penyebaran COVID-19 dengan menentukan wilayah mana yang memiliki tingkat kerentanan menjadi suatu informasi baru bagi institusi kesehatan untuk mencegah penyebaran.

Clustering metode yang biasa digunakan untuk memetakan atau mengklasifikasikan suatu objek adalah algoritma K-Means[8].pada penelitian lainnya menggunakan metode k-mean clustering dapat digunakan untuk menentukan indikator yang mempengaruhi penyebaran covid-19 [8][13].Sebuah penelitian yang dilakukan di amerika serikat yaitu GIS-COVID19 mengelompokkan 3 cluster ,Mereka adalah cluster wilayah dengan tingkat pendapat yang tidak sama ,umr yang rendah dan partisipasi perawat. Faktor risiko ini berpengaruh signifikan dalam hasil COVID-19[9].

Penelitian lainnya yang menggunakan metode k-mean clustering untuk mengidentifikasi faktor memungkinkan terinfeksi COVID-19.bawa Kami mengidentifikasi beberapa indikator atau tanda utama kemungkinan terjadi pada beberapa populasi terinfeksi[13]. Misalnya Diare, mual, konjungtivitis dan hilangnya rasa ditemukan berada pada tahap paling awal dari infeksi. Juga, riwayat perjalanan seperti flu gejala bisa menjadi indikator kuat dari infeksi[13].

hasilnya orang tua, pasien Tidak Sehat (Komorbiditas) dengan Hipertensi, Diabetes, kriteria Covid bersama dengan indikator utama berkembangnya virus selama seminggu, Orang Muda, Sehat dengan kriteria Covid berkembang selama 14 hari. Usia tengah , memiliki gangguan kesehatan (Komorbiditas), Hipertensi, Penyakit jantung, kriteria Covid tanpa indicator berkembangnya virus selama 1 minggu[13].

Dapat disimpulkan Memahami pola Covid yaitu indikator penyebabnya merupakan subjek penelitian yang berkelanjutan. Meskipun, ada beberapa set aturan atau petunjuk yang dapat menunjukkan adanya terkena virus, unsupervised learning seperti clustering bisa menjadi analisis yang kuat dalam meningkatkan kualitas kesehatan[13].

1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini pihak satgas covid perlu menentukan faktor yang berpengaruh terhadap tingkat kasus covid-19. disamping itu faktor komorbiditas dan populasi penduduk berpengaruh terhadap tingginya kasus covid-19, banyaknya variasi variable ini menyebabkan proses pembentukan cluster tidak sederhana sehingga diperlukan sistem untuk menentukan pengaruh dari variable yang mana terhadap tingkatan kasus covid.

1.3. Batasan Masalah

Pada Penelitian ini terdapat Batasan/Ruang Lingkup permasalahan yang telah ditentukan seperti :

- a. Data yang dianalisa adalah data komorbid dan jumlah penduduk berlokasi di seluruh kecamatan kota bandung

- b. Data lokasi seluruh kecamatan kota bandung diambil dari repository publik dengan format geojson.
- c. Data yang diambil adalah hasil rekapitulasi profil kesehatan tahun 2020 dari dinas kesehatan kota bandung.
- d. Variable yang dilakukan penelitian adalah obesitas,diabetes,hipertensi,tuberculosis,diare ,jumlah penduduk dan covid.

1.4. Tujuan penelitian

Terdapat tujuan penelitian dengan adanya penelitian ini dan dilaksanakannya penelitian sebagai berikut:

- a. Menemukan cluster dengan karakteristik tinggi atau rendahnya kasus covid-19 dan melihat seberapa pengaruh transmisi covid terhadap kasus covid sehingga mitigasi dapat dilakukan untuk mengurangi persebaran kasus covid.
- b. Hasil dari clustering akan divisualisi dalam bentuk map untuk melihat wilayah mana yang termasuk pada suatu cluster.

1.5. Luaran dan Manfaat Penelitian

Luaran dari penelitian ini adalah sistem yang dapat mengelompokan wilayah sesuai dengan karakteristik clusternya dan menunjukkan cluster mana yang memiliki pengaruh tinggi terhadap kasus covid dan transmisi covid.

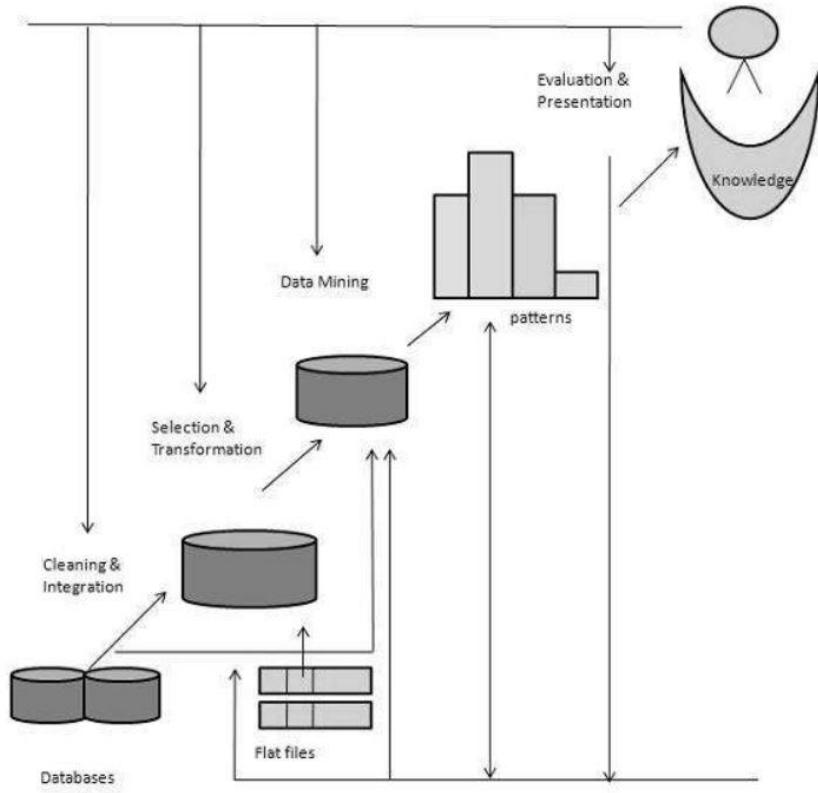
:

Manfaat dari penelitian ini adalah diharapakan dapat menjadi referensi untuk dilakukannya penanggulangan covid dan transmisi covid pada wilayah kota bandung.

1.6. Metode Penelitian

Knowledge Discovery in Database (KDD) merupakan proses penemuan pengetahuan dalam database. Secara lengkap KDD didefinisikan sebagai proses ekstraksi atau identifikasi pola, pengetahuan dan informasi potensial

dari sekumpulan data yang besar. Pengetahuan dan informasi yang dihasilkan dari KDD bersifat sah, baru, mudah dimengerti, dan bermanfaat[31].



1.1 Gambar Knowledge Discovery Data

1. Data selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-processing / cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

3. Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. Interpretation / evalution

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

1.7. Sistematik Penelitian

Sistematik penulisan memperjelas ruang lingkup pembahasan, penyusunan laporan penelitian tugas akhir dibuat dengan sistematika yang terdiri dari:

BAB I	<p>PENDAHULUAN</p> <p>Bab I berisi uraian mengenai latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, metode penelitian, luaran penelitian dan sistematika penulisan.</p>
-------	--

BAB II	TINJAUAN PUSTAKA Bab 2 berisi studi literatur ,landasan teori dan variable penelitian
BAB III	ANALISIS DAN PERANCANGAN SISTEM Bab 3 Analisis sistem berjalan,perancangan sistem,preprocessing data, metode elbow,k-mean clustering, rata-rata variable,perancangan perangkat lunak
BAB IV	IMPLEMENTASI DAN PENGUJIAN SISTEM Bab 4 berisi Implementasi sistem , pengujian perangkat lunak dan pengujian cluster
BAB V	KESIMPULAN DAN SARAN

BAB II TINJAUAN PUSTAKA

Tinjauan pustaka merupakan penelitian terdahulu yang memiliki kaitan dengan penelitian yang dilakukan serta review teori penunjang dalam penelitian, sebagai berikut:

2.1. Studi Literatur

Terdapat studi literatur yang merupakan proses mengumpulkan informasi yang diperoleh dari tesis sebelumnya dengan hasil evaluasi , sehingga dapat diketahui rekomendasi apa yang perlu ditempuh dari penelitian sebelumnya ataupun pengetahuan baru dengan variable yang berbeda.

2.1.1. Pengelompokan provinsi terdampak pandemi covid-19 di Indonesia
Penelitian ini menggunakan analisis deskriptif dan analisis cluster dengan metode K-Means dari data kasus pandemi Covid-19 setiap Provinsi yang ada di Indonesia yaitu berupa data provinsi, kasus pasien positif, kasus sembuh, kasus meninggal, ODP, dan PDP[20]. hasilnya diperoleh 6 kelompok dari analisis kluster dan dapat diberi tingkatan berdasarkan hasil profilisasi tingkat 1 (paling tinggi) ke 6 (paling rendah)[20]Kesimpulan yang didapat adalah bahwa daerah di Indonesia yang memiliki kerentanan dalam menghadapi wabah virus Covid-19 yaitu kota-kota besar seperti Provinsi DKI Jakarta dan Jawa Barat[20].

2.1.2. Correlation Between Temperature and COVID-19 (Suspected, Confirmed and Death)

Pada penelitian ini menggunakan metode k-mean clustering dengan data kasus covid-19 yang didapat dari WHO berupa populasi , jumlah kematian , jumlah terkonfirmasi,jumlah yang dicuriga , tinggi temperature dan rendah temperature pada distrik yang ada di china[21]terlihat kota hubei ini memiliki jumlah kasus yang dicurigai

tertinggi , dibandingkan dengan kota-kota lain. Singkatnya, dari analisis data, kami menemukan bahwa suhu tidak hanya faktor signifikan dalam penyebaran pandemi COVID-19. Kitadapat mengatakan bahwa atribut lain juga menjadi pengaruh dalam COVID-19[21].

2.1.3. Identification of spatial variations in COVID-19 epidemiological data using K-Means clustering algorithm

kami telah menggunakan metode K-means clustering pada negara teridentifikasi covid-19 dengan data epidemiologi dan latar belakang yang secara karakteristik berpengaruh.Tampak jelas bahwa Amerika, negara-negara Eropa, dan Australia termasuk kedalam cluster 2 dengan tingkat kematian kasus COVID-19 yang tinggi, diakibatkan kan oleh faktor proporsi yang lebih tinggi ,populasi negara, jumlah dinyatakan positif COVID-19, persentase PDB yang dibelanjakan lebih tinggi untuk kesehatan masyarakat ,pengeluaran, dan persentase penduduk yang berumur lebih besar dari 65 tahun[22].

2.2. Landasan Teori

Landasan Teori sebagai pernyataan yang disusun secara sistematis dan memiliki variable yang kuat , yang memuat hasil-hasil penelitian yang dimana teori dan hasil penelitian yang digunakan sebagai teori untuk menyelesaikan studi kasus/permashalan.

2.2.1. K-Mean Clustering

K-Means adalah metode pengelompokan yang cepat dan sederhana dengan jumlah iterasi yang lebih kecil. Algoritma ini membagi data menjadi bagian k. Persyaratan cluster diperkirakan berdasarkan pengguna pilihan. Komputer secara acak memilih dan menetapkan objek ke satu objek cluster (k). Jarak antara setiap objek dan pusat setiap cluster dihitung dan menghasilkan cluster yang optimal larutan[14].

Cara kerja algoritma k-Means adalah sebagai berikut:

1. Tentukan jumlah cluster K.
2. Inisialisasi centroid secara acak dan akan memilih titik data k untuk centroid tanpa penggantian.
3. Terus lakukan iterasi sampai tidak ada perubahan pada centroid, misalnya penungasa poin data ke klustes tidak berubah
4. Hitung jumlah jarak kuadrat antara titik data dan semua centroid

$$d_{ij} = \sqrt{\sum_{k=1}^P \{X_{ik} - X_{jk}\}^2}$$

(2)euclidian distance

Keterangan :

Dij : Jarak objek antara objek i dan j

P : Dimensi data

Xik : Koordinat dari obyek i pada dimensi k

Xjk : Koordinat dari obyek j pada dimensi

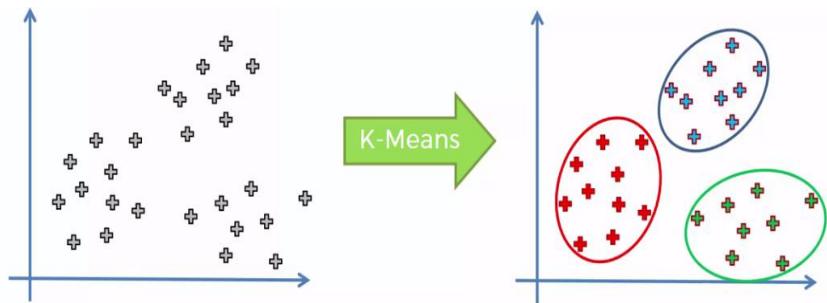
5. Tetapkan setiap titik data ke cluster terdekat
6. Hitung centroid untuk cluster dengan mengambil rata-rata dari semua titik data yang dimiliki setiap cluster.

$$\mu_k = \frac{1}{N_k} \sum_{j=1}^{N_k} x_j$$

(3) Rata-rata centroid

Keterangan :

- μ_k :adalah titik tengah kelompok ke-I/k
 Nk :banyak data dalam cluster ke-I/k
 I,K :indeks dari cluster
 J :indeks dari variable(v1,v2,v3..)
 Xj :nilai data ke-j dalam ke-I/k



2.1 Gambar K-mean Clustering

Hasil pengelompokan data point sesuai dengan gambar(), sehingga setiap data point berkumpul sesuai dengan centroid terdekatnya, sampai titik centroid tidak berpindah lagi sehingga dapat disimpulkan cluster tetapnya.

2.2.2. Komparasi Metode Clustering

Pada Studi komparasi yang dilakukan oleh peneliti lain bahwa algoritma ini[15]. Pengelompokan tertinggi dan terendah Akurasinya masing-masing adalah (94,7%, 77,1%) dan (94,4%, 88,5%) . Rata-rata keseluruhan positif akurasi prediksi yang diperoleh dengan pendekatan ini sekitar 92%. Dalam kasus FCM, akurasi pengelompokan tertinggi dan terendah adalah (97,2%, 91,1%), (97,2%, 90,9%), (97,8%, 90,4%), dan (97,1%, 90,2%)[15].sehingga menghasilkan Dalam hal waktu

komputasi algoritma k-means jauh lebih baik dari algoritma FCM. Sedangkan waktu komputasi tinggi dalam algoritma FCM[15].menemukan bahwa k-means, cenderung menghasilkan akurasi tinggi di semua percobaan. Di sisi lain, pengelompokan hierarkis menyajikan lebih banyak hal kinerja yang terbatas dalam mengelompokkan kumpulan data besar, menghasilkan akurasi yang rendah dalam beberapa eksperimen[16].pada penelitian yang menggunakan data komorbid disimpulkan bahwa k-mean lebih baik digunakan karena waktu komputasi dan akurasi yang tinggi.

2.2.3. Elbow Method

Metode ini berfokus pada persentase varian sebagai fungsi jumlah cluster.Nilai dari k ditambahkan satu per satu dan nilai Sum Square Error (SSE) direkam[14].

$$SSE = \sum_{K=1}^K \sum_{xi \in Sk} \|Xi - Ck\|_2^2$$

(4) Sum of Square Error

Keterangan :

SSE : Sum of Square Error

Xi : nilai atribut dari data ke-i

Ck : nilai atribut titik pusat cluster ke-I

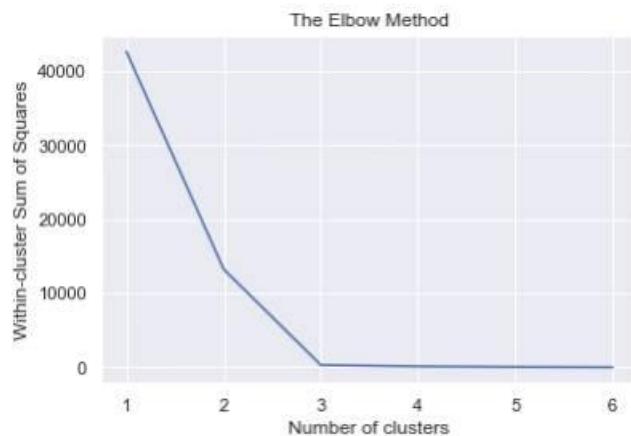
K : Banyaknya cluster

metode elbow ini digunakan untuk menentukan jumlah k optimal dari jumlah data yang ada. Cara kerja algoritma elbow adalah sebagai berikut :

1. awal nilai cluster
2. Menaikkan nilai cluster sampai dengan jumlah cluster

3. Menghitung nilai SSE(Sum of Squares Error)
4. Melakukan perhitungan SSE dengan jumlah cluster yang telah ditentukan
5. Melihat hasil SSE (Sum of Squares Error) dari nilai cluster yang turun secara drastis atau nilai yang berubah signifikan.
6. Menetapkan nilai cluster yang berbentuk siku.

Mulai dari $k = 2$ dan nilai SSE kemudian ditambahkan langkah dengan langkah, di mana $k_n = k + 1$, SSE_{k_n} terbesar - SSE_{k_n-1} adalah titik di mana nilai k optimal ditemukan. Ketika nilainya dari k ditambahkan kembali maka cluster baru tersebut mirip dengan cluster sebelumnya cluster atau jumlah kesalahan tidak berubah secara signifikan yang menghasilkan nilai k [14].



2.2 Gambar Elbow method

Saat nilainya turun secara drastis dan membentuk sudut yang lebih kecil(membentuk seperti siku-siku tangan), maka nilai k optimalnya

telah ditemukan. semakin banyak jumlah datanya maka tingkat SSEnya semakin tinggi.

2.2.4. Silhouette Coefficient

Metode silhouette coefficient berfungsi untuk menguji kualitas dari cluster yang dihasilkan serta sebagai metode pengesahan suatu cluster yang menggabungkan antara metode cohesion dan metode separation [32]. Untuk perhitungan nilai silhouette coefficient dibutuhkan nilai jarak antar objek dengan menggunakan metode euclidean distance. Tahapan dalam menentukan nilai silhouette coefficient adalah sebagai berikut :

1. Pada setiap objek i dihitung nilai rata-rata titik satu dengan semua objek yang berada dalam satu cluster. Maka akan diperoleh nilai rata-rata yang disebut dengan a_i .
2. Pada setiap objek i dihitung nilai minimal jarak rata-rata dari titik satu ke titik yang lain yang berbeda cluster. Maka akan diperoleh nilai rata-rata minimum yang disebut dengan b_i .
3. Kemudian setelah semua nilai diketahui maka nilai silhouette coefficient dapat ditentukan dengan menggunakan rumus sebagai berikut :

$$Si = \frac{b_i - a_i}{\max(a_i - b_i)}$$

Keterangan :

Dimana : Si : nilai silhouette coefficient

a_i : rata-rata jarak titik cluster n dengan semua data yang berada dalam cluster n

b_i : minimal jarak rata-rata dari titik cluster n ke titik cluster yang lain yang berbeda dari cluster n

Hasil perhitungan nilai silhouette coefficient memiliki range antara -1 hingga 1. Hasil dapat dikatakan baik apabila bernilai positif, hal ini berarti titik sudah berada di dalam cluster yang tepat sedangkan jika nilainya negatif ini menandakan terjadinya overlapping sehingga titik berada di antara dua cluster [33]. Nilai silhouette berdasarkan teori Kaufman dan Rousseeuw :

1.Strong Structure

$0.7 < SC \leq 1$ disebut sebagai cluster yang kuat (sangat baik).

2.Medium Structure

$0.5 < SC \leq 0.7$ disebut sebagai cluster dengan struktur medium.

3.Weak Structure

$0.25 < SC \leq 0.5$ disebut sebagai cluster yang buruk.

4.Nostructure

$SC \leq 0.25$ disebut sebagai cluster yang tidak memiliki struktur.

2.2.5. Min-Max Scaler

Normalisasi min-max melakukan perubahan linier pada data asli. Nilai dinormalisasi dalam rentang yang diberikan semisalnya dari 0 sampai

1. Manfaat normalisasi Min-Max adalah bahwa semua nilai adalah dalam kisaran tertentu[26].

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

(4)Min-Max

Keterangan :

Min : minimum

Max :maximum

X : value range

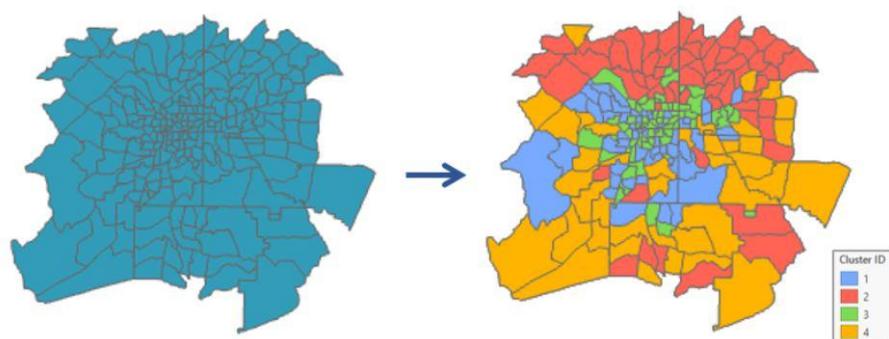
2.2.6. Sistem Informasi Geografis

Sistem Informasi Geografis(GIS) merupakan sebuah tool yang digunakan untuk memetakan wilayah secara geografis , bisa juga digunakan untuk menampilkan tingginya prevalansi ,transmisi suatu penyakit dan memodelkan lingkungan secara spatial[23].

Pada penggunaannya Data yang diolah melalui sistem informasi geografis menampilkan hasil kajian informasi dalam bentuk pemetaan berdasarkan lokasi yang akan ditampilkan[24].

GIS memungkinkan peneliti untuk memasukkan, mengintegrasikan, menganalisis, memodelkan, menafsirkan, dan memvisualisasikan data spasial dengan cara yang mengungkapkan hubungan, pola, dan tren dalam bentuk peta, laporan, dan bagan. Pemetaan membantu menjawab pertanyaan, memecahkan masalah, dan bahkan memasarkan ide penelitian baru dengan mengamati data spasial dengan cara yang mudah dipahami dan dibagikan[25].

Saat ini penggunaan teknologi GIS sangat berkembang dan banyak digunakan untuk memetakan berbagai kasus , hasilnya dapat dilakukan analisa bagaimana pengaruh suatu wilayah terhadap data yang digunakan.



2.3 Gambar Clustering Map

Dengan adanya sistem informasi geografis dapat melakukan visualiasi wilayah yang asalnya belum teridentifikasi clusternya menjadi dikelompokan masing-masing wilayah berdasarkan karakteristik yang sama sehingga dapat menjadi informasi mengenai wilayah yang terkelompok berdasarkan clusternya.

2.2.7. Rata-Rata Variable

Nilai rata-rata adalah angka representasi atau biasa juga disebut sebagai mean dari suatu kelompok data yang mewakili data secara keseluruhan. Dapat digunakan untuk membandingkan kualitas suatu kelompok dengan kelompok lain ,Mewakili kondisi suatu kelompok dengan 1 angka dan Mendeskripsikan suatu kelompok dengan singkat[29].

$$x = \frac{1}{n} \sum_{i=1}^n x_i$$

(5) rata-rata variable

Keterangan :

X = rata-rata hitung

X_i = nilai sample ke-I

N = jumlah sample

2.2.8. Unified Model Language

Unified Modeling Language (UML) adalah bahasa pemodelan standar untuk pengembangan perangkat lunak dan sistem. UML digunakan untuk visualisasi, merancang dan mendokumentasikan sistem piranti lunak. UML menawarkan sebuah standar untuk merancang model sebuah sistem. Dalam UML terdapat beberapa diagram yang digunakan untuk merancang sebuah sistem diantaranya

Use Case Diagram, Class Diagram, State Diagram, Activity Diagram, Sequence Diagram, Component Diagram dan Deployment Diagram.

Pada penelitian terdahulu UML digunakan untuk perancangan sistem penjualan komputer [27]. Melalui pemodelan, persyaratan sistem secara bertahap disempurnakan dan ditingkatkan secara iteratif. Model dapat dengan jelas memandu pengembangan sistem.

2.2.7.1. Use Case Diagram

Use case diagram menggambarkan fungsionalitas yang diharapkan dari sebuah sistem. Sebuah use case merepresentasikan sebuah interaksi antara aktor dengan sistem [27]. Use case digunakan untuk mengetahui fungsi apa saja yang ada di dalam sistem dan siapa saja yang berhak menggunakan fungsi-fungsi tersebut. Dalam use case terdapat seorang atau sebuah aktor adalah sebuah entitas manusia atau mesin yang berinteraksi dengan sistem untuk melakukan pekerjaan-pekerjaan tertentu.

2.2.7.2. Scenario Use Case

Skenario Use Case merupakan sebuah alur yang menjelaskan proses pada setiap Use Case yang terdapat pada Use Case Diagram. Skenario Use Case terbagi menjadi tiga bagian, yang pertama merupakan identifikasi dan inisiasi, kemudian yang kedua merupakan step perfomed dan yang terakhir merupakan kondisi, asumsi dan pertanyaan. Dalam setiap Use Case memiliki skenario normal yang dimana merupakan suatu skenario apabila sistem berjalan normal tanpa terjadi kesalahan atau error. Dan juga skenario alternatif yang dimana skenario apabila sistem tidak berjalan normal, atau mengalami error.

2.2.7.3. Activity Diagram

Activity Diagram menggambarkan berbagai alur aktivitas dalam sistem yang berfokus pada aliran kontrol dari satu aktivitas ke aktivitas lainnya yang dimana decision dapat juga terjadi. Activity diagram juga dapat menggambarkan proses paralel yang mungkin terjadi dan pada prosesnya didorong oleh pemrosesan internal[27].

2.2.7.4. Sequence Diagram

Sequence Diagram menggambarkan interaksi antar objek di dalam dan di sekitar sistem berupa pesan yang digambarkan terhadap waktu. Sequence diagram terdiri atas dimensi vertikal (waktu) dan dimensi horizontal (objek-objek yang terkait). Tujuan utama dari Sequence Diagram adalah untuk mengekspresikan bagaimana objek berinteraksi satu sama lain untuk mengimplementasikan suatu fungsi. Sequence Diagram dapat mewakili model kolaborasi objek yang ada dalam pikiran perancang untuk program pada saat dijalankan [27].

2.2.7.5. Class Diagram

Class Diagram merupakan struktur statis sistem yang digambarkan kelas-kelas mewakili kelas objek dalam sistem dan hubungan antara masing masing kelas, *package* serta objek yang berhubungan satu sama lain [27]. Class Diagram ini berfungsi membantu dalam visualisasi struktur kelas dari suatu sistem.

2.2.7.6. Data Flow Diagram

Data Flow Diagram (DFD) adalah diagram grafis untuk menentukan dan memvisualisasikan model sistem yang akan dikembangkan. DFD berguna dalam mendefinisikan persyaratan pengguna dalam tampilan grafis[28].

2.3. Variable Penelitian

Variable penelitian adalah kriteria yang dimiliki pada suatu objek untuk dilakukan pengamatan dari hasil deskripsi ataupun ilmiah dan kaitannya atas indikator penelitian.

2.3.1. Obesitas

Penumpukan lemak tubuh ini meningkatkan risiko terjadinya gangguan kesehatan serius, seperti penyakit jantung, diabetes, atau hipertensi. Obesitas juga dapat menyebabkan gangguan kualitas hidup dan masalah psikologis. Masyarakat yang di diagnosis memiliki obesitas menjadi pengaruh terinfeksi covid-19 dan membawanya lebih parah[17].

2.3.2. Diabetes

Penyakit diabetes adalah suatu kondisi dimana kadar glukosa atau gula darah meningkat atau di atas batas normal. Glukosa menumpuk di dalam darah akibat tidak diserap sel tubuh dengan baik sehingga dapat menimbulkan berbagai gangguan organ tubuh. Diabetes melitus dikaitkan perkembangan penyakit pada pasien dengan COVID-19[12].

2.3.3. Tuberculosis

Tuberkulosis (TB) Paru adalah penyakit infeksi pada jaringan paru yang disebabkan oleh bakteri *Mycobacterium tuberculosis*. Kuman TB yang terhirup akan masuk kedalam alveoli paru-paru dan mengembangkan lesi kecil yang dinamakan sebagai

fokus primer (fokus Ghon). Selanjutnya infeksi menyebar melalui saluran limfe menuju kelenjar limfe regional. Penyebaran ini mengakibatkan inflamasi di saluran limfe (limfangitis) dan di kelenjar limfe (limfadenitis) yang akan membentuk kompleks primer. Gangguan kesehatan ini dikaitkan dengan Penyakit tuberkulosis dengan peningkatan 2,10 kali lipat risiko penyakit COVID-19 yang parah[10]

2.3.4. Hipertensi

Hipertensi atau penyakit tekanan darah tinggi adalah kondisi dimana tekanan darah dalam arteri meningkat melebihi batas normal. Tekanan adalah menunjukkan tingkat kekuatan dorongan darah pada permukaan pembuluh darah arteri pada saat darah dipompa dari jantung. Orang yang di diagnosa hipertensi memiliki kriteria berkembangnya virus covid-19 selama 1-2 minggu [13]

2.3.5. Diare

Diare adalah penyakit yang ditandai dengan bertambahnya frekuensi defekasi lebih biasanya (> 3 kali/hari) disertai perubahan konsistensi tinja (menjadi cair), dengan/tanpa darah dan/atau lendir. Saat pasien memiliki diare, merupakan gejala yang dapat terkonfirmasi karena pandemi covid-19[18].

2.3.6. Jumlah Penduduk

Jumlah populasi pada suatu kawasan menentukan seberapa aktif mobilitas yang dilakukan sehingga menjadi faktor yang mempengaruhi penyebaran covid-19 ,seperti kepadatan penduduk yang tinggi di kota besar dan daerah metropolitan merupakan faktor risiko utama untuk tertular virus.[19], wilayah kota bandung sendiri

termasuk daerah dengan padat penduduk dan memiliki mobilitas yang tinggi.

2.3.7. Covid-19

Patofisiologi COVID-19 diawali dengan interaksi protein *spike* virus dengan sel manusia. Setelah memasuki sel, *encoding genome* akan terjadi dan memfasilitasi ekspresi gen yang membantu adaptasi virus SARS-CoV-2 (*severe acute respiratory syndrome virus corona 2*) pada inang. Rekombinasi, pertukaran gen, insersi gen, atau delesi, akan menyebabkan perubahan genom yang menyebabkan *outbreak*. Virus-virus yang dikeluarkan kemudian akan menginfeksi sel ginjal, hati, intestinal, dan limfosit T, dan traktus respiratorius bawah, yang kemudian menyebabkan gejala pada pasien. Gejala dan tanda COVID-19 terutama berupa infeksi saluran napas, tetapi dapat juga menyebabkan di saluran pencernaan seperti diare, mual, dan muntah, jantung seperti miokarditis, saraf seperti anosmia bahkan stroke, serta mata dan kulit.

Pada tahap awal, jumlah terinfeksi kasus sedikit meningkat, namun, setelah mencapai 100 kasus yang dikonfirmasi terinfeksi, tingkat pertumbuhannya cepat atau sedikit ada perubahan[4]

BAB III ANALISIS DAN PERANCANGAN

Dalam bab ini berisi mengenai analisis sistem berjalan , perancangan sistem analis kesehatan masyarakat yang rentan terhadap paparan covid-19 serta perancangan perangkat lunak dan perancangan algortima.

3.1. Analisis Sistem Berjalan

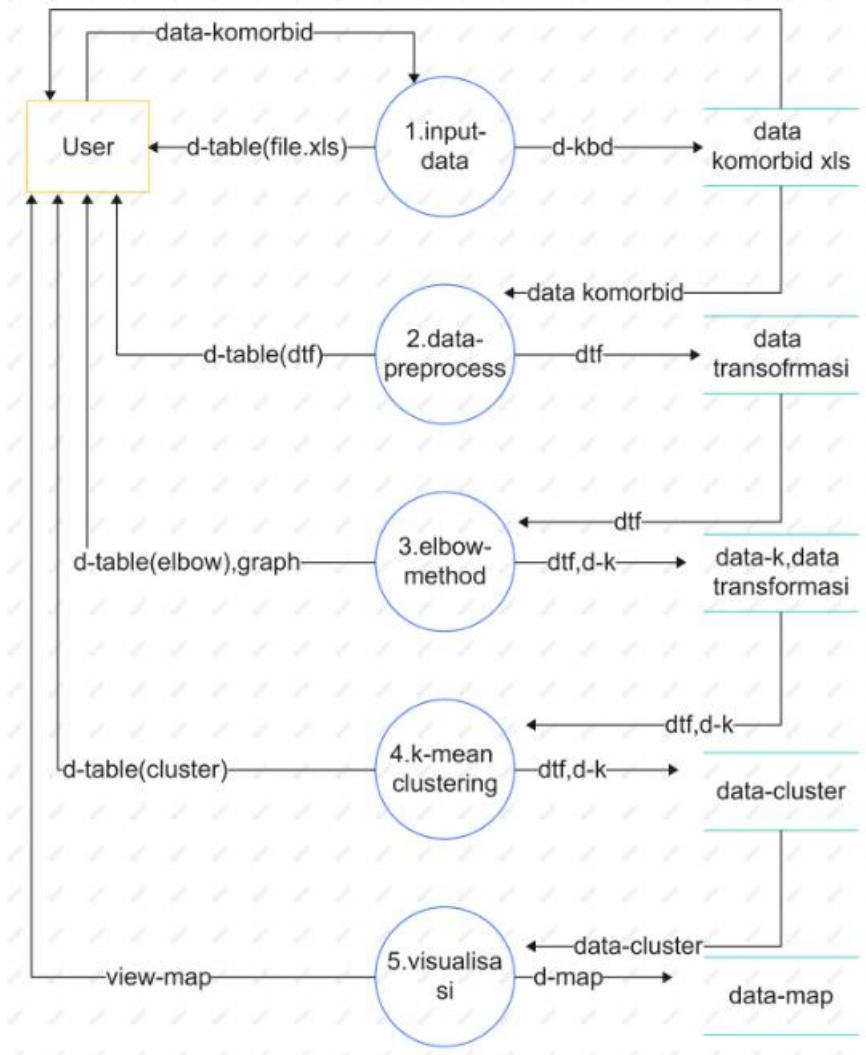
Sistem clustering wilayah yang rentan terhadap paparan covid-19 merupakan sebuah sistem yang mampu melakukan pengelompokan terhadap data inputan yang sudah diproses sehingga cluster wilayah dapat ditentukan ,penerapan sistem clustering wilayah yang rentan terhadap paparan covid-19 ini digunakan untuk membantu instansi kesehatan dan masyarakat sehingga dapat mengetahui wilayah mana yang memiliki tingkat kerentan tinggi ataupun rendah.

Analisis cluster wilayah ini merupakan sebuah proses untuk mengevaluasi wilayah-wilayah yang belum teridentifikasi akan clusternya , mentransformasi data mentah dan menggunakan data tersebut sehingga wilayah akan dikelompokan berdasarkan karakteristik wilayah yang sama.

3.2. Perancangan Sistem Analisis kesehatan masyarakat yang rentan terhadap paparan covid-19 menggunakan metode k-mean clustering dan visualisasi sig

Pada sistem Analisis kesehatan masyarakat yang rentan terhadap paparan covid-19 ini digambarkan dengan data flow diagram yang dimana menggunakan teknik data mining sebagai metodenya, untuk data yang digunakan adalah data komorbid/penyakit penyerta yang berelasi dengan covid-19 ,data ini berjumlah 30 wilayah dengan 6 variable , untuk prosesnya sistem ini akan membaca file yang diimport oleh pengguna setelah itu lanjut ke tahap pre-process, dari data yang sudah dipreprocess akan menghasilkan data yang siap digunakan untuk melakukan cluster wilayah. dalam proses elbow method akan menghasilkan jumlah cluster optimal yang digunakan untuk clusternya, sampai ke tahap k-mean selesai didapat hasil untuk dievaluasi dan di visualisasi.

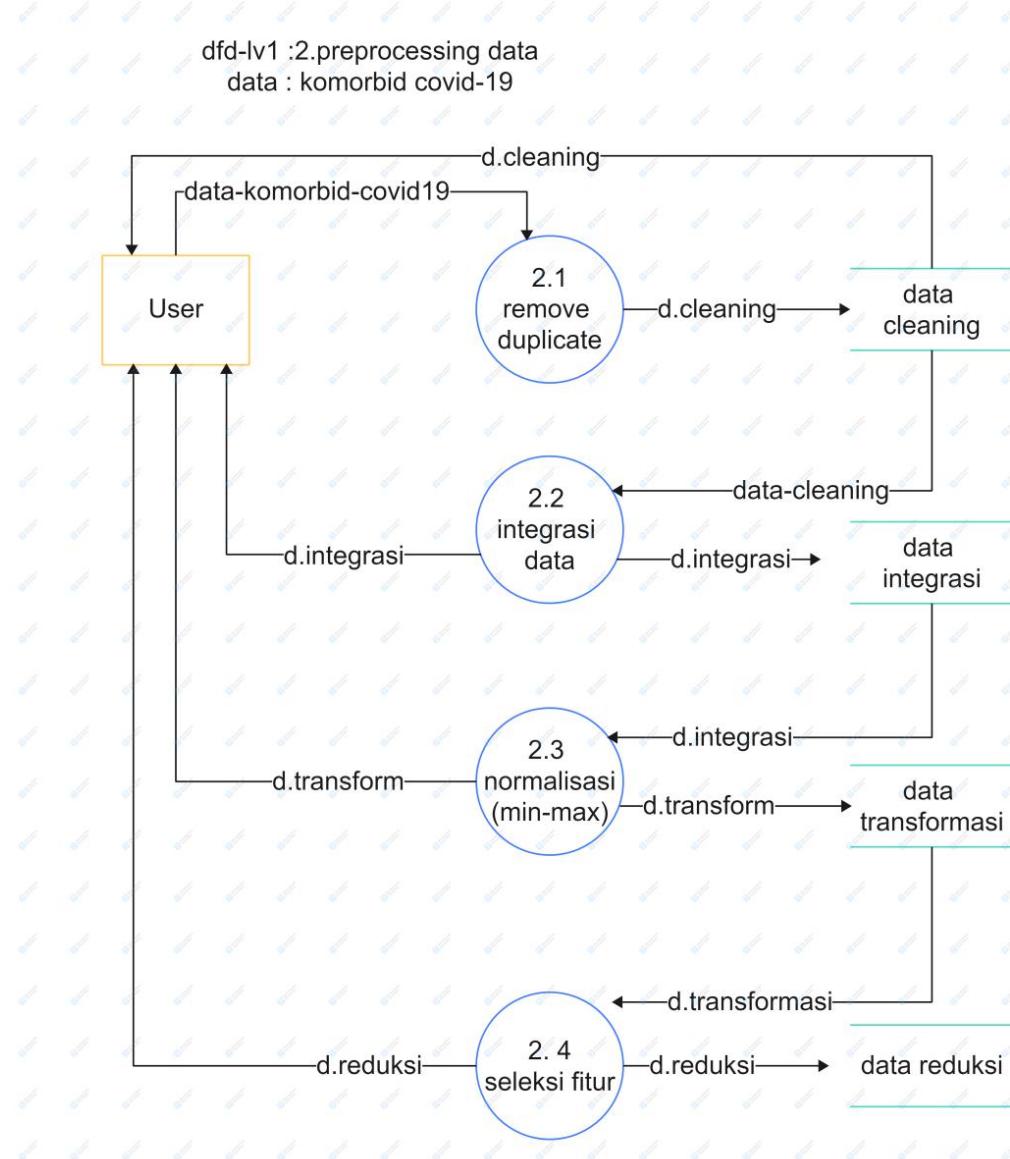
dfd lv0 ,perancangan sistem dengan teknik data mining



3.1 Gambar Data Flow Diagram level 0

3.3. Pre-processing Data komorbid covid-19

Dalam preprocessing data adalah tahapan sistem untuk mengelola data sehingga data itu siap itu digunakan , digambarkan dengan data flow diagram untuk mengetahui bagaimana data itu diproses dan apa hasil dari proses tersebut.



3.2 Gambar Data Flow Diagram level 1

3.3.1. Remove Duplicate

pada tahapan pertama apabila terdapat variable yang duplikasi akan dihilangkan atau digabungkan jumlahnya untuk menghindari kemiripan variable.

3.1 Table Dengan Variable Duplikasi

no	Kecamatan	Pender	Pender	Diabet	TB_BT	TB_	Diare	Hipert	Jumla	posi
----	-----------	--------	--------	--------	-------	-----	-------	--------	-------	------

		ita_Ob esitas_	ita_Ob esitas_	es_Me	A_L	BTA _P		ensi	h_Pen	tif_c
		L	P	llitus				duduk	ovid	
1	Andir	107	447	1130	48	40	1607	4004	99378	297
2	Antapani	89	190	1164	10	6	2801	4751	78817	301
3	Arcamanik	23	154	1918	14	11	2159	5387	76913	316
4	Astanaany ar	815	1116	499	16	21	688	2774	74049	186
5	Babakan Ciparay	473	1070	429	57	35	2023	4481	13937 4	181
6	Bandung Kidul	0	0	462	14	17	1318	2023	60129	134
7	Bandung Kulon	114	243	829	34	30	1612	2364	13327 5	353
8	Bandung Wetan	101	369	242	9	4	722	1260	29019	44
9	Batunungg al	496	1270	2493	28	11	1397	7628	12132 4	153
10	Bojongloa Kaler	904	982	584	38	27	1903	3856	12356 9	153
11	Bojongloa Kidul	193	369	878	37	26	2674	4352	86004	185
12	Buahbatu	0	0	758	14	13	874	4447	10142 6	223
13	Cibeunying Kaler	465	510	630	14	2	2712	2114	69917	156
14	Cibeunying Kidul	0	0	1006	25	20	4192	7076	11311 0	184
15	Cibiru	34	586	317	13	18	2223	1383	73114	160
16	Cicendo	1505	7095	963	18	12	2359	2355	96413	218
17	Cidadap	0	0	453	21	14	1056	2126	53960	75
18	Cinambo	45	172	102	6	1	1712	627	25266	54
19	Coblong	0	28	1053	35	34	2351	5443	11466	336

									5	
20	Gedebage	0	132	759	7	1	814	3196	40398	109
21	Kiaracondong	192	1250	330	28	26	2390	4812	130347	244
22	Lengkong	91	160	1049	16	12	688	4861	71448	183
23	Mandalajati	52	366	386	12	11	1749	2792	71403	133
24	Panyileukan	0	0	302	7	5	1492	777	39634	162
25	Rancasari	397	642	487	8	11	657	2628	83904	190
26	Regol	16	73	909	14	20	1903	3497	80515	208
27	Sukajadi	16	185	1369	18	23	1779	7985	101740	269
28	Sukasari	7	231	875	16	19	2298	4283	77284	147
29	Sumur Bandung	116	411	234	5	9	1141	4926	37187	115
30	Ujungberung	176	784	386	24	15	1889	1418	86804	176

Dari data table 3.1, terdapat kolom yang akan dieleminasi dikarenakan memiliki variable yang sama perbedaanya terdapat pada jenis kelamin, seperti variable TB dan Obesitas . Sehingga hasil eliminasi dan penggabungan didapat 6 variable.

3. 2 Table Setelah Dikakukan Elimnasi Variable

no	Kecamatan	Obesitas	Diabetes	Tuberculosis	Diare	Hipertensi	Jumlah_Penduduk	positif_covid
1	Andir	554	1130	88	1607	4004	99378	297

2	Antapani	279	1164	16	2801	4751	78817	301
3	Arcamanik	177	1918	25	2159	5387	76913	316
4	Astanaanyar	1931	499	37	688	2774	74049	186
5	Babakan Ciparay	1543	429	92	2023	4481	139374	181
6	Bandung Kidul	0	462	31	1318	2023	60129	134
7	Bandung Kulon	357	829	64	1612	2364	133275	353
8	Bandung Wetan	470	242	13	722	1260	29019	44
9	Batununggal	1766	2493	39	1397	7628	121324	153
10	Bojongloa Kaler	1886	584	65	1903	3856	123569	153
11	Bojongloa Kidul	562	878	63	2674	4352	86004	185
12	Buahbatu	0	758	27	874	4447	101426	223
13	Cibeunying Kaler	975	630	16	2712	2114	69917	156
14	Cibeunying Kidul	0	1006	45	4192	7076	113110	184
15	Cibiru	620	317	31	2223	1383	73114	160
16	Cicendo	8600	963	30	2359	2355	96413	218
17	Cidadap	0	453	35	1056	2126	53960	75
18	Cinambo	217	102	7	1712	627	25266	54
19	Coblong	28	1053	69	2351	5443	114665	336
20	Gedebage	132	759	8	814	3196	40398	109
21	Kiaracondon	1442	330	54	2390	4812	130347	244
22	Lengkong	251	1049	28	688	4861	71448	183

23	Mandalajati	418	386	23	1749	2792	71403	133
24	Panyileukan	0	302	12	1492	777	39634	162
25	Rancasari	1039	487	19	657	2628	83904	190
26	Regol	89	909	34	1903	3497	80515	208
27	Sukajadi	201	1369	41	1779	7985	101740	269
28	Sukasari	238	875	35	2298	4283	77284	147
29	Sumur Bandung	527	234	14	1141	4926	37187	115
30	Ujungberung	960	386	39	1889	1418	86804	176

3.3.2. Data Integration

Pada tahapan kedua ini akan menggabungkan data dari source yang terpisah agar sesuai dengan format yang dapat diperhitungkan , semisalnya akan menggabungkan 2 file yaitu table hipertensi.xls dan table diabetes.xls yang didapat dari source yang berbeda.

3. 3 Table Dengan Variable Terpisah

no	Kecamatan	Diabetes		no	Kecamatan	Hipertensi
1	Andir	1130		1	Andir	4004
2	Antapani	1164		2	Antapani	4751
3	Arcamanik	1918		3	Arcamanik	5387
4	Astanaanyar	499		4	Astanaanyar	2774
5	Babakan Ciparay	429		5	Babakan Ciparay	4481
6	Bandung Kidul	462		6	Bandung Kidul	2023

7	Bandung Kulon	829		7	Bandung Kulon	2364
8	Bandung Wetan	242		8	Bandung Wetan	1260
9	Batununggal	2493		9	Batununggal	7628
10	Bojongloa Kaler	584		10	Bojongloa Kaler	3856
11	Bojongloa Kidul	878		11	Bojongloa Kidul	4352
12	Buahbatu	758		12	Buahbatu	4447
13	Cibeunying Kaler	630		13	Cibeunying Kaler	2114
14	Cibeunying Kidul	1006		14	Cibeunying Kidul	7076
15	Cibiru	317		15	Cibiru	1383
16	Cicendo	963		16	Cicendo	2355
17	Cidadap	453		17	Cidadap	2126
18	Cinambo	102		18	Cinambo	627
19	Coblong	1053		19	Coblong	5443
‘20	Gedebage	759		‘20	Gedebage	3196
21	Kiaracondong	330		21	Kiaracondong	4812
22	Lengkong	1049		22	Lengkong	4861
23	Mandalajati	386		23	Mandalajati	2792
24	Panyileukan	302		24	Panyileukan	777
25	Rancasari	487		25	Rancasari	2628
26	Regol	909		26	Regol	3497
27	Sukajadi	1369		27	Sukajadi	7985

28	Sukasari	875		28	Sukasari	4283
29	Sumur Bandung	234		29	Sumur Bandung	4926
30	Ujungberung	386		30	Ujungberung	1418

Dikarenakan setiap file memiliki variable wilayah yang sama maka akan dilakukan eliminasi data objek yang sama ,setelah itu menggabungkan 2 kolom sehingga hasilnya seperti pada table 3.4.

3. 4 Table yang telah dilakukan Data Integrasi

no	Kecamatan	Diabetes	Hipertensi
1	Andir	1130	4004
2	Antapani	1164	4751
3	Arcamanik	1918	5387
4	Astanaanyar	499	2774
5	Babakan Ciparay	429	4481
6	Bandung Kidul	462	2023
7	Bandung Kulon	829	2364
8	Bandung Wetan	242	1260
9	Batununggal	2493	7628
10	Bojongloa Kaler	584	3856
11	Bojongloa Kidul	878	4352
12	Buahbatu	758	4447
13	Cibeunying Kaler	630	2114
14	Cibeunying Kidul	1006	7076
15	Cibiru	317	1383
16	Cicendo	963	2355

17	Cidadap	453	2126
18	Cinambo	102	627
19	Coblong	1053	5443
20	Gedebage	759	3196
21	Kiaracondong	330	4812
22	Lengkong	1049	4861
23	Mandalajati	386	2792
24	Panyileukan	302	777
25	Rancasari	487	2628
26	Regol	909	3497
27	Sukajadi	1369	7985
28	Sukasari	875	4283
29	Sumur Bandung	234	4926
30	Ujungberung	386	1418

3.3.3. Min-Max Scaler

Pada tahapan ketiga ini melakukan Normalisasi min-max pada data asli. Nilai dinormalisasi dalam rentang yang diberikan semisalnya dari 0 sampai 2. Manfaat normalisasi Min-Max adalah bahwa semua nilai adalah dalam kisaran tertentu.dari hasil normalisasi dengan interval 0 sampai 2 ,didapat data point dengan nilai minimal terendah 0 ,nilai medium 1 dan nilai maksimal 2.

3. 5 Table Min-Max Scaler interval 0-2

Kecamatan	Obesitas	Diabetes	Tuberculosis	Diare	Hipertensi	Jumlah_Penduduk	positif_covid

Andir	0.128 83720 9	0.859891 259	1.905 8823 53	0.537482 32	0.9179124 76	1.2989799 14	1.637540 453
Antapani	0.064 88372 1	0.888331 242	0.211 7647 06	1.213012 73	1.1209567 82	0.9386020 26	1.663430 421
Arcamanik	0.041 16279 1	1.519029 695	0.423 5294 12	0.849787 836	1.2938298 45	0.9052301 33	1.760517 799
Astanaanyar	0.449 06976 7	0.332078 628	0.705 8823 53	0.017538 897	0.5835824 95	0.8550320 75	0.919093 851
Babakan Ciparay	0.358 83720 9	0.273525 721	2	0.772842 999	1.0475672 74	2	0.886731 392
Bandung Kidul	0	0.301129 235	0.564 7058 82	0.373974 54	0.3794509 38	0.6110526 87	0.582524 272
Bandung Kulon	0.083 02325 6	0.608113 76	1.341 1764 71	0.540311 174	0.4721391 68	1.8931012 72	2
Bandung Wetan	0.109 30232 6	0.117105 813	0.141 1764 71	0.036775 106	0.1720576 24	0.0657797 88	0
Batununggal	0.410 69767 4	2	0.752 9411 76	0.418670 438	1.9029627 62	1.6836330 49	0.705501 618
Bojongloa Kaler	0.438 60465 1	0.403178 586	1.364 7058 82	0.704950 495	0.8776841 53	1.7229817 37	0.705501 618
Bojongloa Kidul	0.130 69767 4	0.649100 795	1.317 6470 59	1.141159 83	1.0125033 98	1.0645704 07	0.912621 359
Buahbatu	0	0.548724 383	0.470 5882 35	0.122772 277	1.0383256 32	1.3348757 32	1.158576 052
Cibeunying Kaler	0.226 74418	0.441656 211	0.211 7647	1.162659 123	0.4041859 2	0.7826094 58	0.724919 094

	6		06					
Cibeunying Kidul	0	0.756168 967	0.894 1176 47	2	1.7529219 9	1.5396641 78	0.906148 867	
Cibiru	0.144 18604 7	0.179841 071	0.564 7058 82	0.885997 171	0.2054906 22	0.8386440 92	0.750809 061	
Cicendo	2	0.720200 753	0.541 1764 71	0.962942 008	0.4696928 51	1.2470116 03	1.126213 592	
Cidadap	0	0.293601 004	0.658 8235 29	0.225742 574	0.4074476 76	0.5029270 52	0.200647 249	
Cinambo	0.050 46511 6	0	0	0.596888 26	0	0	0	0.064724 919
Coblong	0.006 51162 8	0.795483 061	1.458 8235 29	0.958415 842	1.3090513 73	1.5669190 59	1.889967 638	
Gedebage	0.030 69767 4	0.549560 853	0.023 5294 12	0.088826 025	0.6982875 78	0.2652224 21	0.420711 974	
Kiaracondong	0.335 34883 7	0.190715 182	1.105 8823 53	0.980480 905	1.1375373 74	1.8417814 7	1.294498 382	
Lengkong	0.058 37209 3	0.792137 181	0.494 1176 47	0.017538 897	1.1508562 11	0.8094436 85	0.899676 375	
Mandalajati	0.097 20930 2	0.237557 507	0.376 4705 88	0.617821 782	0.5884751 29	0.8086549 58	0.576051 78	
Panyileukan	0	0.167294 019	0.117 6470 59	0.472418 67	0.0407719 49	0.2518315 98	0.763754 045	
Rancasari	0.241 62790 7	0.322040 987	0.282 3529 41	0	0.5438977 98	1.0277631 72	0.944983 819	

Regol	0.020 69767 4	0.675031 368	0.635 2941 18	0.704950 495	0.7801032 89	0.9683633 05	1.061488 673
Sukajadi	0.046 74418 6	1.059807 612	0.8	0.634794 908	2	1.3403792 9	1.456310 68
Sukasari	0.055 34883 7	0.646591 384	0.658 8235 29	0.928429 986	0.9937483 01	0.9117327 44	0.666666 667
Sumur Bandung	0.122 55814	0.110414 053	0.164 7058 82	0.273833 098	1.1685240 55	0.2089424 05	0.459546 926
Ujungberung	0.223 25581 4	0.237557 507	0.752 9411 76	0.697029 703	0.2150040 77	1.0785922 11	0.854368 932

3.3.4. Feature Selection

Pada tahapan terakhir ini akan seleksi atau mengeliminasi variable yang tidak memiliki relasi dengan covid-19 berdasarkan hasil dari penelitian publik yang berkaitan dengan data komorbid covid-19. berikut adalah variable yang dieleminasi.

3. 6 Table Feature Selection Variable yang tidak digunakan

Kecamatan	Penyakit jiwa	Demam berdarah
Andir	125	66
Antapani	120	79
Arcamanik	81	139
Astanaanyar	143	66
Babakan Ciparay	103	90
Bandung Kidul	135	65

Bandung Kulon	53	64
Bandung Wetan	58	45
Batununggal	96	99
Bojongloa Kaler	95	93
Bojongloa Kidul	118	50
Buahbatu	148	180
Cibeunying Kaler	96	85
Cibeunying Kidul	105	90
Cibiru	70	72
Cicendo	121	77
Cidadap	87	60
Cinambo	142	36
Coblong	164	139
Gedebage	156	64
Kiaracondong	82	111
Lengkong	69	84
Mandalajati	64	95
Panyileukan	31	76
Rancasari	82	153
Regol	75	40
Sukajadi	41	48
Sukasari	157	100
Sumur Bandung	131	59
Ujungberung	93	134

3.4. Metode Elbow

Metode elbow adalah metode untuk menentukan jumlah cluster yang optimal dari data point yang dimiliki yaitu data komorbid, untuk menentukannya dengan cara melihat hasil selisih/ sum of square error dari setiap cluster, apabila cluster memiliki jumlah penurunan sum of square error yang signifikan maka dapat ditentukan jumlah cluster yang optimalnya.

Tahap 1: tentukan jumlah cluster , jumlah cluster ditentukan sebanyak 5, yang dimana perulangan dimulai dari jumlah K - 1.tentukan titik pusat cluster secara random.

Centro id	Obesitas	diabetes	tuberculosis	diare	hipertensi	Jumlah_penduduk	Covid-19
Centro id 1	0.2575	0.7871	0.9926	0.8419	1.1329	1.3911	1.2394
Centro id 2	0.1252	0.2915	0.3613	0.3905	0.4684	0.5790	0.5829

Tahap 2 : melakukan perhitungan SSE

$$SSE = \sum_{k=1}^K \sum_{xi \in Sk} \|Xi - Ck\|_2^2$$

2.4 Gambar Sum of Square Error

Keterangan :

SSE : Sum of Square Error

Xi : nilai atribut dari data ke-I

Ck : nilai atribut titik pusat cluster ke-I

K : Banyaknya cluster

Data ke-i	Perhitungan SSE -i (xi-ck)	Penjumlahan ke-i
$\sum_{2=1}^2 \sum_{x1 \in 2} \ xi - ck\ _2^2$	(0.128837209 - 0.2575872091875)^2 (0.859891259 - 0.7871183605)^2 (1.905882353 - 0.9926470588125)^2 (0.53748232 - 0.8419377651875)^2 (0.917912476 - 1.13293354175)^2 (1.298979914 - 1.3911141199375)^2 (1.637540453 - 1.2394822006875)^2	1.1617374208125
$\sum_{2=1}^2 \sum_{x2 \in 2} \ xi - ck\ _2^2$	(0.064883721 - 0.2575872091875)^2 (0.888331242 - 0.7871183605)^2 (0.211764706 - 0.9926470588125)^2 (1.21301273 - 0.8419377651875)^2 (1.120956782 - 1.13293354175)^2 (0.938602026 - 1.3911141199375)^2 (1.663430421 - 1.2394822006875)^2	1.1794952916257
.	.	.
.	.	.
$\sum_{2=1}^2 \sum_{x30 \in 2} \ xi - ck\ _2^2$	(0.223255814 - 0.12524916942857)^2 (0.237557507 - 0.29156957635714)^2 (0.752941176 - 0.36134453771429)^2 (0.697029703 - 0.39050313185714)^2 (0.215004077 - 0.46843086228571)^2 (1.078592211 - 0.57903540014286)^2	0.6472595888386 3
		Hasil penjumlahan

total SSE dari
data ke-1 sampai
ke-30 didapat
untuk cluster 2 =
29.357775054012

Tahap 3 : Menaikan jumlah cluster , perhitungan SSE ditambah 1 cluster sampai dengan jumlah cluster yang telah ditentukan. Lalu menentukan centroid yang baru.

Centro id	Obesitas	diabetes	tuberculosis	diare	hipertensi	Jumlah_penduduk	Covid-19
Centro id 1	0.1800	0.8286	1.2149	0.8671	1.2476	1.5324	1.2868
Centro id 2	0.3418	0.5422	0.5058	0.8347	0.6534	1.0125	0.5829
Centro id 3	0.1053	0.2929	0.3208	0.2473	0.5212	0.4915	1.0008

Tahap 4 : Melakukan Perhitungan SSE kembali, dengan jumlah centroid baru 3.

Data ke-i	Perhitungan SSE -i ($x_i - c_k$)	Penjumlahan ke-i

$\sum_{3=1}^3 \sum_{x1 \in 3} \ xi - ck\ _2^2$	(0.649100795 - 0.82863769436364)^2 (1.317647059 - 1.214973262)^2 (1.14115983 - 0.86717243154545)^2 (1.012503398 - 1.2476463466364)^2 (1.064570407 - 1.5324764099091)^2 (0.912621359 - 1.2868490732727)^2 (0.130697674 - 0.18004228318182)^2	0.53455400777226
$\sum_{3=1}^3 \sum_{x2 \in 3} \ xi - ck\ _2^2$	(0 - 0.341889534875)^2 (0.548724383 - 0.542241739875)^2 (0.470588235 - 0.505882352875)^2 (0.122772277 - 0.834724186625)^2 (1.038325632 - 0.65343843425)^2 (1.334875732 - 1.012553896375)^2 (1.158576052 - 1.0008090615)^2	0.90197161909813
.	.	.
.	.	.
$\sum_{3=1}^3 \sum_{x30 \in 3} \ xi - ck\ _2^2$	(0.223255814 - 0.341889534875)^2 (0.237557507 - 0.542241739875)^2 (0.752941176 - 0.505882352875)^2 (0.697029703 - 0.834724186625)^2 (0.215004077 - 0.65343843425)^2 (1.078592211 - 1.012553896375)^2 (0.854368932 - 1.0008090615)^2	0.40493473053965
	Hasil penjumlahan total SSE dari data ke-1 sampai ke-30 Didapat untuk cluster 3 =25.676783332866	

Tahap 5 : lakukan perhitungan sampai dengan jumlah cluster ke-5, sehingga didapat hasil SSE tiap cluster dan selisihnya dapat ditetapkan

jumlah cluster yang optimal adalah sebanyak 3 dengan 30 data dan 6 variable dari data komorbid , karena perpindahan dari cluster 2 ke cluster 3 memiliki penurunan seliyang signifikan.

cluster	SSE	Selisih (SSE i - SSE -i+1)
2	29.35	29.35
3	25.67	3.680
4	21.36	4.308
5	22.07	0.704

3.5. K-mean Clustering

K-Means adalah metode pengelompokan data yang bertujuan data point yang memiliki karakteristik sama akan ditentukan pada kelompok yang mana , cara kerja k-mean untuk pengelompokan data yaitu menggunakan rumus euclidean distance melihat kedekatan data dengan centroid random.

Tahap 1 : Menentukan jumlah K/cluster , pada tahapan ini jumlah cluster ditentukan dari hasil metode elbow yang didapat jumlah k optimal sebanyak 3 cluster.

Tahap 2 : Menentukan Centroid secara acak sebanyak dengan jumlah clusternya. Sehingga dapat dilakukan perhitungan kedekatan data point dengan centroidnya.

Centro id	Obesitas	diabetes	tuberculos is	diar e	hipertensi	Jumlah_pend uduk	positi fcovid
1	0.358837	0.2735257	2	0.77	1.04756727	2	0.8867

	209	21		2842 999	4		31392
2	0 67	0.7561689 7	0.89411764	2	1.75292199	1.539664178	0.9061 48867
3	0.058372 093	0.7921371 81	0.49411764 7	0.01 7538 897	1.15085621 1	1.150856211	0.8996 76375

Tahap 3 : Setelah diketahui centroid untuk masing-masing variable, lalu masuk ke proses perhitungan euclidean distance untuk menentukan suatu data jarak paling dekatnya dengan centroid yang mana.

$$d_{ij} = \sqrt{\sum_{k=1}^P \{X_{ik} - X_{jk}\}^2}$$

(2)

Keterangan :

Dij : Jarak objek antara objek i dan j

P : Dimensi data

X_{ik} : Koordinat dari obyek i pada dimensi k

X_{jk} : Koordinat dari obyek j pada dimensi

Data ke-1 Andir	Proses Euclidean distance	C1	C2	C3
$d_{1,1} = \sqrt{\sum_{k=1}^6 \{x_{ik} - x_{jk}\}^2}$	$\sqrt{(0.128837209 - 0.358837209)^2 + (0.859891259 - 0.273525721)^2 + (1.905882353 - 2)^2 + (0.53748232 - 0.772842999)^2}$	1.238 1159 4713		

Data 1 , centroid 1	$+(0.917912476 - 1.047567274)^2$ $+(1.298979914 - 2)^2$ $+(1.637540453 - 0.886731392)^2$	49		
$d_{1,2} \sqrt{\sum_{k=1}^6 \{x_{ik} - x_{jk}\}^2}$ Data 1 , centroid 2	$\sqrt{(0.128837209 - 0)^2}$ $+ (0.859891259 - 0.756168967)^2$ $+ (1.905882353 - 0.894117647)^2$ $+ (0.53748232 - 0.017538897)^2$ $+ (0.917912476 - 1.75292199)^2$ $+ (1.298979914 - 1.539664178)^2$ $+ (1.637540453 - 0.906148867)^2$	2.116 6215 0144 51		
$d_{1,3} \sqrt{\sum_{k=1}^6 \{x_{ik} - x_{jk}\}^2}$ Data 1 , centroid 3	$\sqrt{(0.128837209 - 0.058372093)^2}$ $+ (0.859891259 - 0.792137181)^2$ $+ (1.905882353 - 0.494117647)^2$ $+ (0.53748232 - 0.928429986)^2$ $+ (0.917912476 - 0.017538897)^2$ $+ (1.298979914 - 1.150856211)^2$ $+ (1.637540453 - 0.899676375)^2$	1.763 8958 5765 96		

Tahap 4 : Penentuan cluster yang pertama , dari perhitungan euclidean distance didapat jarak terdeket antara objek dengan centroidnya , setelah itu dapat dilihat untuk wilayah andir temasuk ke cluster 1 karena dari ketiga centroid yang paling kecil jaraknya adalah C1.

no	Objek	Cluster
1	Andir	1
2	Antapani	3
3	Arcamanik	3

4	Astanaanyar	3
5	Babakan Ciparay	1
6	Bandung Kidul	3
7	Bandung Kulon	1
8	Bandung Wetan	3
9	Batununggal	3
10	Bojongloa Kaler	1
11	Bojongloa Kidul	1
12	Buahbatu	3
13	Cibeunying Kaler	3
14	Cibeunying Kidul	2
..		
..		
30	Ujung Berung	3

Tahap 5 : Penentuan Centroid baru pertama,penentuan centroid ini dilakukan agar menyesuaikan data dengan centroid yang sebenarnya.Dilakukan penentuan centroid sampai dengan tidak ada perubahan centroid .

$$\mu_k = \frac{1}{N_k} \sum_{j=1}^{N_k} x_j$$

(6)Rata-Rata Centroid

Keterangan :

μ_k : adalah titik tengah kelompok ke-I/k

N_k : banyak data dalam cluster ke-I/k

I,K : indeks dari cluster

J : indeks dari variable(v₁,v₂,v₃..)

X_j : nilai data ke-j dalam ke-I/k

Centroid	Penentan centroid
$\mu_1 = \frac{1}{5} \sum_{j=1}^{N_k} x_j$ Centroid 1	$\left(\frac{1.637540453+2+0.886731392+0.705501618+0.912621359}{5} \right)$ $\approx 0.26424418575, (\text{obesitas})$ $\left(\frac{0.859891259+0.273525721+0.60811376+0.403178586+0.649100795}{5} \right)$ $\approx 0.54642409025, (\text{diabetes})$ $\left(\frac{1.905882353+1.341176471+2+0.403178586+0.649100795}{5} \right)$ $\approx 1.6470588235 (\text{tuberculosis})$ $\left(\frac{0.53748232+0.772842999+0.540311174+0.704950495+1.14115983}{5} \right)$ $\approx 0.789108911 (\text{diare})$ $\left(\frac{0.917912476+1.047567274+0.472139168+0.877684153+1.012503398}{5} \right)$ $\approx 0.96391682525 (\text{hipertensi})$ $\left(\frac{1.298979914+2+0.611052687+1.722981737+1.064570407+1.893101272}{5} \right),$ $\approx 1.5216330145 (\text{jumlah penduduk})$ $\left(\frac{1.637540453+2+0.886731392+0.705501618+0.912621359}{5} \right)$ $\approx 1.0355987055 (\text{positif covid})$

Tahap 6 : Setelah centroid baru ditemukan, lakukan lagi langkah untuk menghitung euclidean distance dengan centroid yang baru sampai penentuan cluster menjadi tetap sehingga didapat cluster tetapnya ,

no	Kecamatan	Cluster
----	-----------	---------

1	Andir	1
2	Antapani	2
3	Arcamanik	1
4	Astanaanyar	3
5	Babakan Ciparay	1
6	Bandung Kidul	3
7	Bandung Kulon	1
8	Bandung Wetan	3
9	Batununggal	1
10	Bojongloa Kaler	1
11	Bojongloa Kidul	1
12	Buahbatu	2
13	Cibeunying Kaler	2
14	Cibeunying Kidul	1
15	Cibiru	2
16	Cicendo	2
17	Cidadap	3
18	Cinambo	3
19	Coblong	1
20	Gedebage	3
21	Kiaracondong	1
22	Lengkong	3
23	Mandalajati	3
24	Panyileukan	3
25	Rancasari	3
26	Regol	2
27	Sukajadi	1
28	Sukasari	2

29	Sumur Bandung	3
30	Ujungberung	2

3.6. Rata-Rata Variable

Nilai rata-rata adalah angka representasi atau biasa juga disebut sebagai mean dari suatu kelompok data yang mewakili data secara keseluruhan. Dapat digunakan untuk membandingkan kualitas suatu kelompok dengan kelompok lain ,Mewakili kondisi suatu kelompok dengan 1 angka dan Mendeskripsikan suatu kelompok dengan singkat.

$$x = \frac{1}{n} \sum_{i=1}^n x_i$$

(7) Rata-Rata variable

Keterangan :

X = rata-rata hitung

X_i = nilai sample ke-I

N = jumlah sample

Perhitungan	Proses Perhitungan
$X = \frac{1}{11} \sum_{i=1}^n 1_i$ Nilai sample 1	$\frac{1}{11}(554+177+1543+357+1766+1886+562+0+28+1442+201)$ ≈ 774
$X = \frac{1}{8} \sum_{i=1}^n 1_i$ Nilai sample 2	$\frac{1}{8}(279+0+975+620+8600+89+238+960)$ ≈ 1470
$X = \frac{1}{11} \sum_{i=1}^n 1_i$ Nilai sample 3	$\frac{1}{11}(1931+0+470+0+217+132+251+418+0+1039+527)$ ≈ 453

Berdasarkan perhitungan yang dihasilkan dari masing-masing variable, didapatkan karakteristik pada masing-masing cluster.

cluster	obesitas	diabetes	Tuberculosis	diare	hipertensi	Jumlah_penduduk	Positif covid
1	774	1092	58	2189	5217	112699	242
2	1470	750	28.5	2132	3031	83036	198
3	453	452	20	1094	2544	53308	125

berikut adalah penjelasan untuk setiap cluster.

- Cluster 1 , memiliki karakteristik nilai rata-rata variable diabetes , tuberculosis,diare,hipertensi , jumlah penduduk dan angka positif covid yang lebih tinggi dari cluster 2 dan 3
- Cluster 2 , memiliki karakteristik nilai rata-rata variable,obesitas yang lebih tinggi dari cluster 1 dan 3
- Cluster 3 , memiliki karakteristik nilai rata-rata variable obesitas ,diabetes,tuberculosis,diare,hipertensi,jumlah penduduk dan positif covid lebih kecil dari cluster 2 dan 1.

3.7. Perancangan Perangkat Lunak

Perancangan perangkat lunak yang dibutuhkan untuk menggambarkan sistem clustering wilayah rentan paparan covid-19 yang dibangun dengan mengimplementasikan ke dalam kode program. Perancangan perangkat lunak ini terdiri dari Analisis Kebutuhan Perangkat Lunak, Business Use Case, Use Case Diagram, Skenario Use Case, Activity Diagram, Statechart Diagram, Analisis Dataset, Perancangan Antarmuka, dan Perancangan Algoritma.

3.4.1. Fungsional Perangkat Lunak

Perangkat lunak yang dibutuhkan merupakan perangkat lunak yang dapat melakukan clustering wilayah yang rentan akan paparan covid-19. oleh karena itu, fungsi fungsi perangkat lunak yang dibutuhkan dideskripsikan pada bagian kebutuhan fungsional. Adapun rincian kebutuhan fungsional perangkat lunak sebagai fungsi utama yang di antaranya :

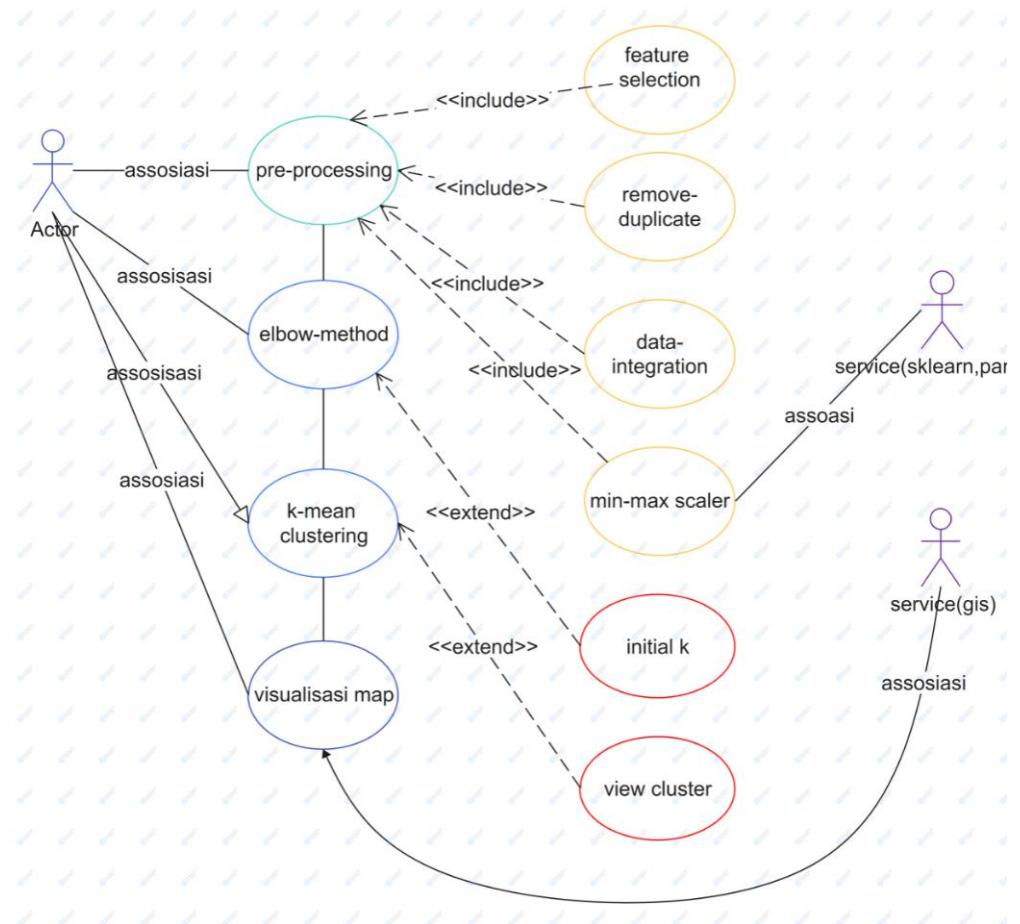
3. 7 Table Fungsional Perangkat Lunak

no	Kebutuhan	Aktor	Keterangan
	Fungsional		
1	Pre-processing	User,service(sklearn,pandas) dari python library	Fungsional sistem yang melakukan pemrosesan data numerik untuk dilakukan cleaning ,normalisasi dan reduksi
2	Elbow method	User	Fungsional sistem yang dimana user dapat melakukan input jumlah cluster dari rentang yang akan ditentukan, untuk mengukur jumlah cluster yang optimal
3	K-mean Clustering	User	Fungsional sistem untuk menentukan suatu wilayah

			teridentifikasi pada cluster yang mana/mengelompokan data point sesuai dengan centroid yang dekat.
4	Visualisasi Map	User,Service(gis)	Fungsional sistem untuk memetakan data point yang sudah memiliki cluster ke dalam map lalu mewarnai cluster tersebut sesuai dengan karakteristik data point yang dimiliki.

3.4.2. Use Case Diagram

Use Case Diagram untuk sistem analisis clustering wilayah ini memiliki 4 modul utama yaitu preprocessing, elbow method,k-mean dan visualiasi dan use case diagram dapat dilihat pada gambar 3.3.



3.1 Gambar Use Case Diagram

3.4.3. Skenario Use Case

Skenario Use Case dari sistem analisis cluster wilayah ini terdiri dari preprocessing,elbow method,k-mean clustering dan visualisasi map

3.4.3.1. Skenario Use Case Pre-processing

Skenario use case ini terdiri dari feature selection,remove duplicate,data integration min-max scaler.

1. Remove Duplicate

Skenario Use Case remove duplicate ini menjelaskan mengenai interaksi antara aksi aktor dan reaksi sistem analisis wilayah cluster yang dapat dilihat pada table 3.8.

3. 8 Table Use Case Scenario Remove Duplicate

No	1
Nama	Remove Duplicate
Tujuan	Menghapus variable yang mirip.
Deskripsi	Suatu tahapan data cleaning yang akan menghapus variable sama ,sehingga hasilnya hanya ada 1 variable.
Aktor	user
Pre-condition	Data masih memiliki duplikasi.
Post-condition	Data sudah tidak memiliki duplikasi
Skenario Utama	
Aksi Aktor	Reaksi Sistem
1.memilih kolom mana yang akan dihapus dan dijumlahkan	

	2.menghapus kolom dan menjumlahkan data
	3.menampilkan hasil setelah diproses

2. Data Integration

Skenario Use Case Data Integration ini menjelaskan mengenai interaksi antara aksi aktor dan reaksi sistem analisis wilayah cluster yang dapat dilihat pada table 3.9.

3. 9 Table Use Case Scenario data integration

No	2
Nama	Data Integration
Tujuan	Menggabungkan 2 file data
Deskripsi	Suatu tahapan data ekstraksi dari 2 file yang berbeda sehingga data tersebut digabung menjadi 1 file.
Aktor	user
Pre-condition	Data masih terpisah
Post-condition	Data sudah digabungkan
Skenario Utama	
Aksi Aktor	Reaksi Sistem
1.aktor memilih 2 jenis file yang sama	

	2.mengabungkan ke 2 file menjadi 1 file
	3.data baru dikembalikan dan dapat dilihat.

3. Min-max Scaler

Skenario Use Case Min-Max Scaler ini menjelaskan mengenai interaksi antara aksi aktor dan reaksi sistem analisis wilayah cluster yang dapat dilihat pada table 3.10.

3. 10 Table Use Case Scenario Min-max scaler

No	3
Nama	Min-max scaler
Tujuan	Normalisasi data
Deskripsi	Suatu tahapan data dilakukan normalisasi menggunakan metode min-max scaler dengan interval 0-2 sehingga rentang datanya tidak terlalu jauh,
Aktor	User,service(pandas,sklearn)
Pre-condition	Rentang data terlalu jauh
Post-condition	Data memiliki interval tertinggi 2 dan terkecil 0

Skenario Utama	
Aksi Aktor	Reaksi Sistem
1.aktor import data xls	
	2.data import ditampilkan
3.menentukan interval minimal dan maksimalnya	
	4.melakukan proses min-max scaler
	5.hasil min-max scaler ditampilkan

4. Feature Selection

Skenario Use Case Feature Selection ini menjelaskan mengenai interaksi antara aksi aktor dan reaksi sistem analisis wilayah cluster yang dapat dilihat pada table 3.11.

3. 11 Table Use Case Scenario Feature Selection

No	4
Nama	Feature Selection
Tujuan	Menghapus variable yang tidak perlu digunakan
Deskripsi	Suatu tahapan data variable dilakukan penghapusan untuk mereduksi data, karena data tersebut tidak memiliki korelasi.
Aktor	User

Pre-condition	Terdapat data yang tidak digunakan
Post-condition	Data telah dihapus
Skenario Utama	
Aksi Aktor	Reaksi Sistem
1.aktor import data xls	
	2.data import ditampilkan
3.menentukan column mana yang akan dihapus	
	4.melakukan proses penghapusan
	5.hasil penghapusan ditampilkan

3.4.3.2. Skenario Use Case Elbow Method

Skenario use case ini terdiri dari elbow method dan initial jumlah k

1. Elbow Method dan initial-k

Skenario Use Case Elbow method ini menjelaskan mengenai interaksi antara aksi aktor dan reaksi sistem analisis wilayah cluster yang dapat dilihat pada table 3.12.

3. 12 Table Use Case Scenario elbow method

No	1
Nama	Elbow method dan initial-k
Tujuan	Menentukan jumlah cluster yang optimal
Deskripsi	Suatu tahapan untuk menentukan

	jumlah cluster yang optimal dengan menggunakan rumus Sum of Square.
Aktor	User
Pre-condition	Jumlah cluster optimal belum teridentifikasi
Post-condition	Sudah teridentifikasi
Skenario Utama	
Aksi Aktor	Reaksi Sistem
1.aktor membuka halaman elbow method	
	2.halaman ditampilkan
3.menentukan jumlah cluster.	
4.menentukan jumlah iterasi	
	5.menampilkan grafik elbow method

3.4.3.3. Scenario Use Case K-mean clustering

Skenario use case k-mean clustering terdiri dari melihat hasil clustering wilayah sesuai dengan jumlah k-optimal.

1. k-mean clustering dan view cluster

Skenario Use Case k-mean clustering ini menjelaskan mengenai interaksi antara aksi aktor dan reaksi sistem analisis wilayah cluster yang dapat dilihat pada table 3.13.

3. 13 Table Use Case Scenario k-mean

No	1
Nama	K-mean Clustering dan view cluster
Tujuan	Melakukan cluster wilayah
Deskripsi	Pada tahapan ini objek akan dilakukan clustering menggunakan metode-kmean clustering sehingga setiap objek memiliki kelompoknya masing-masing.
Aktor	User
Pre-condition	Objek belum tercluster
Post-condition	Objek memiliki cluster
Skenario Utama	
Aksi Aktor	Reaksi Sistem
1. aktor membuka halaman k-mean	
	2.menampilkan halaman
	3.wilayah sudah dilakukan clustering

3.4.3.4. Scenario Use Case Visualiasi Map

Skenario use case k-mean clustering ini terdiri dari visualisasi map dengan service gis.

1. Visualisasi Map

Skenario Use Case Visualisasi Map ini menjelaskan mengenai interaksi antara aksi aktor dan reaksi sistem analisis wilayah cluster yang dapat dilihat pada table 3.14.

3. 14 Table Use Case Scenario visualiasi map

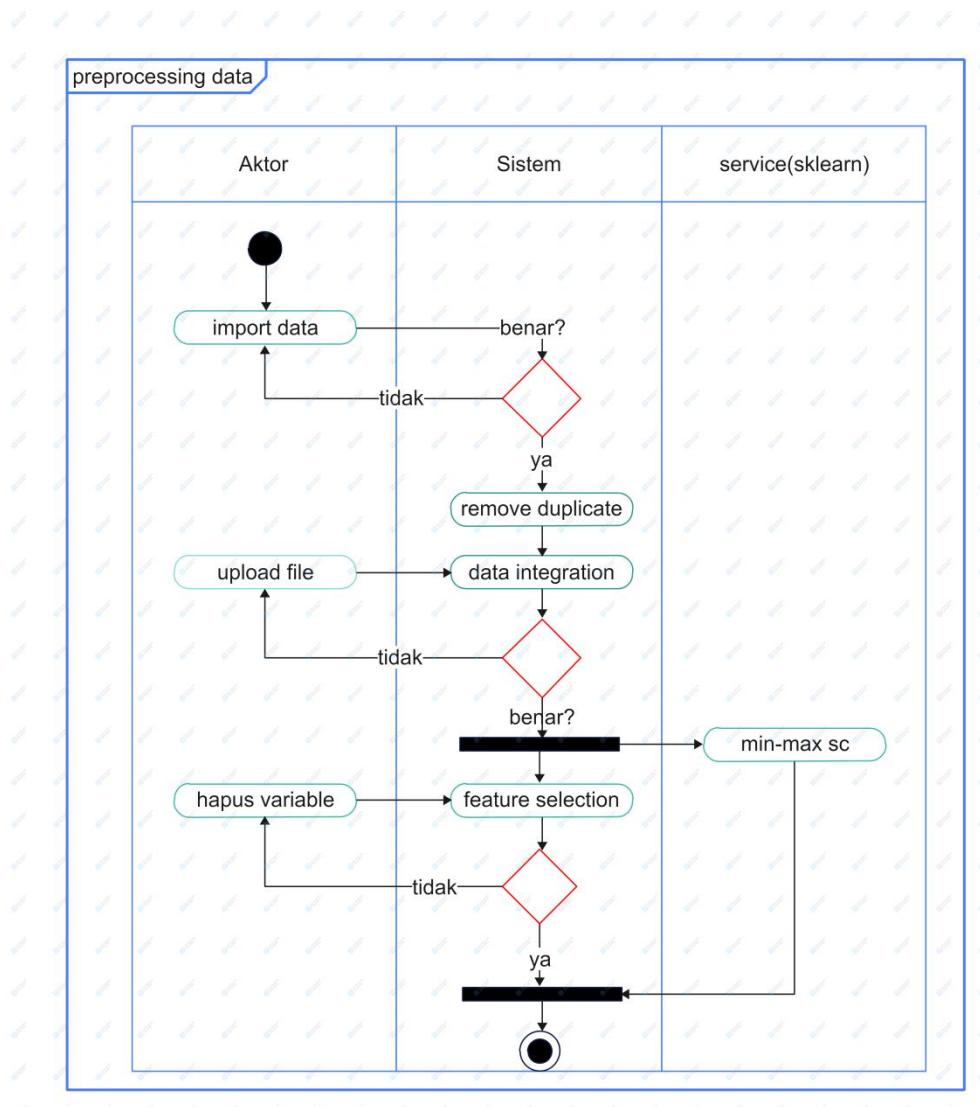
No	1
Nama	Visualiasi map
Tujuan	Melakukan visualiasi map wilayah
Deskripsi	Pada tahapan ini wilayah yang sudah memiliki cluster , akan ditampilkan dalam map sesuai dengan karakteristik clusternya dengan warna.
Aktor	User,gis
Pre-condition	Cluster belum ditampilkan dalam map
Post-condition	Cluster ditampilkan dalam map
Skenario Utama	
Aksi Aktor	Reaksi Sistem
1.aktor membuka halaman gis	
	2.proses cluster sesuai dengan data geojson
	3.menampilkan map

3.4.4. Activity Diagram

Activity Diagram merupakan salah satu tahap pada analisis perancangan perangkat lunak yang menjelaskan urutan aktivitas yang dilakukan oleh pengguna terhadap sistem dan untuk memodelkan sebuah proses yang ada didalam sistem cluster wilayah terdiri dari preprocessing, elbow-method, k-mean dan visualiasi.

3. 4. 1. Activity Diagram Preprocessing

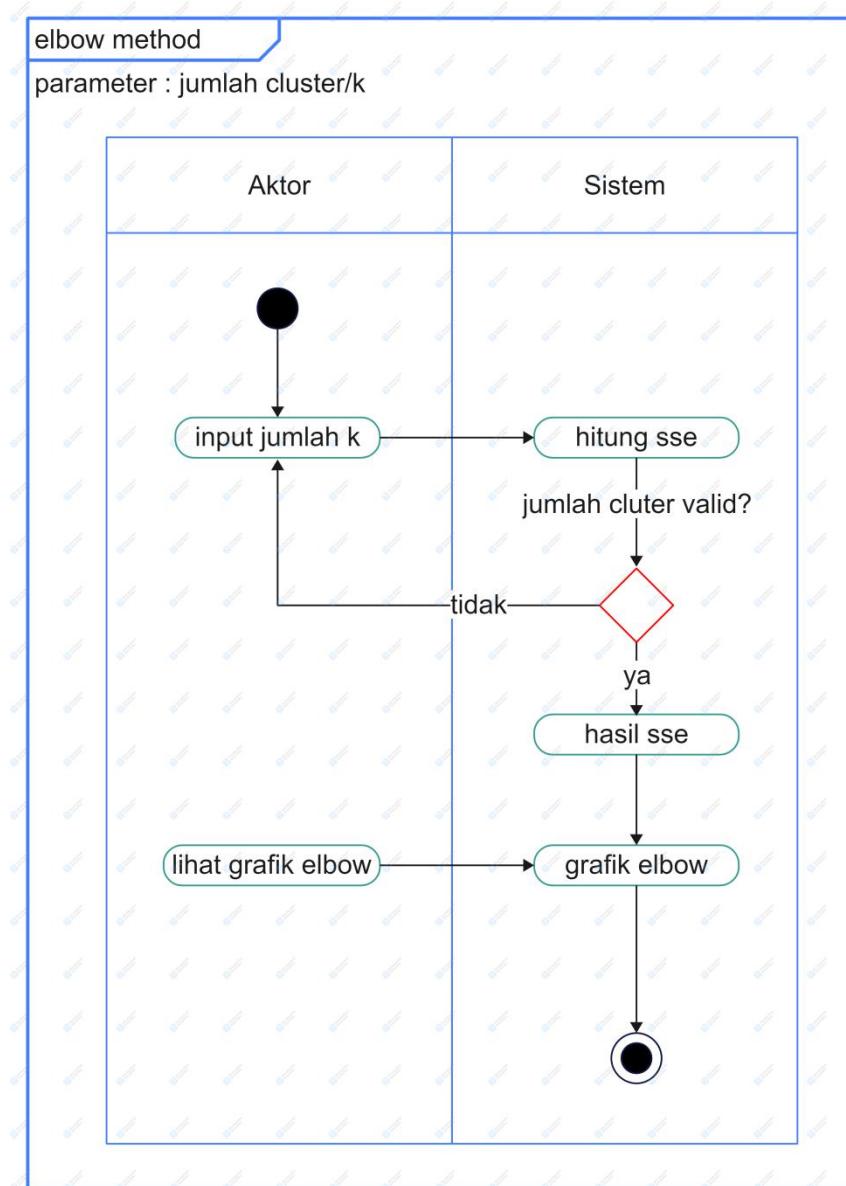
Pada Activity Diagram preprocessing data menggambarkan urutan aktivitas antara sistem terhadap aksi yang dilakukan oleh user untuk melakukan tahapan proses pengolahan data.



3.2 Gambar Activity Diagram preprocessing

3.4.2. Activity Diagram Elbow-method

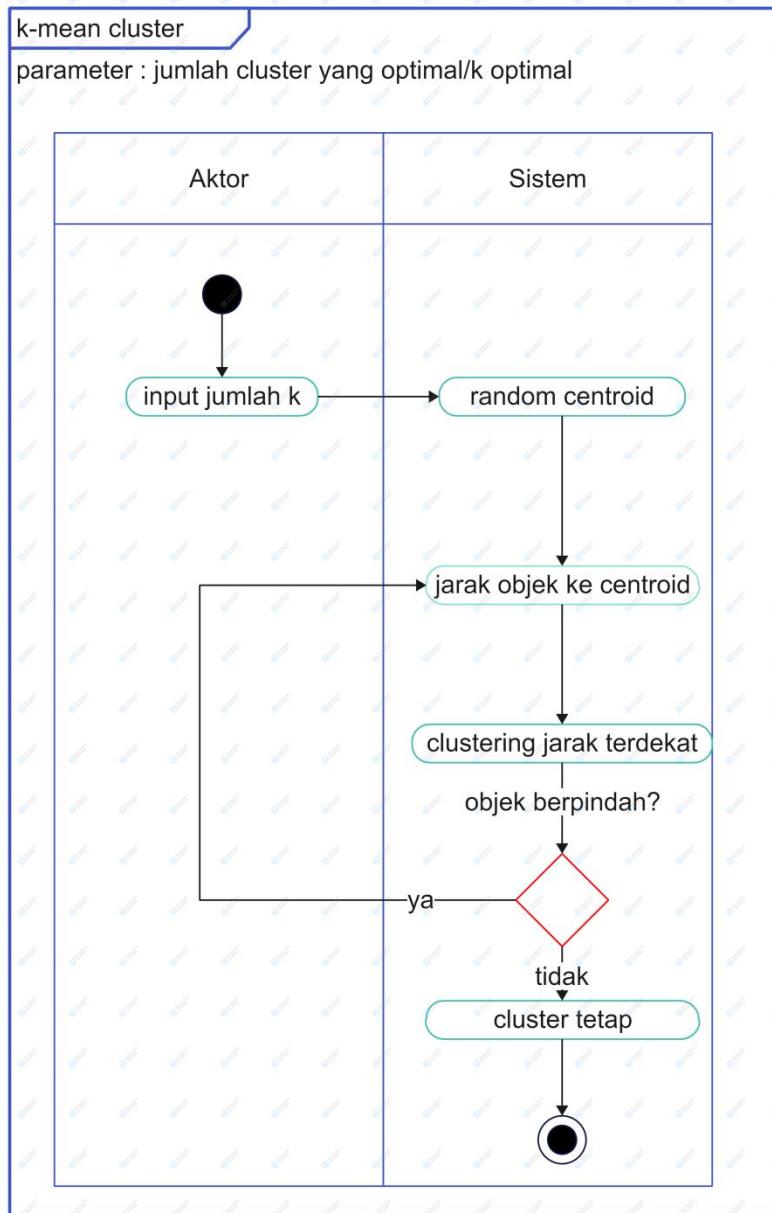
Pada Activity Diagram elbow method menggambarkan urutan aktivitas antara sistem terhadap aksi yang dilakukan oleh user untuk melakukan penentuan jumlah cluster yang optimal.



3.3 Activity Diagram Elbow method

3. 4. 3. Activity Diagram K-mean clustering

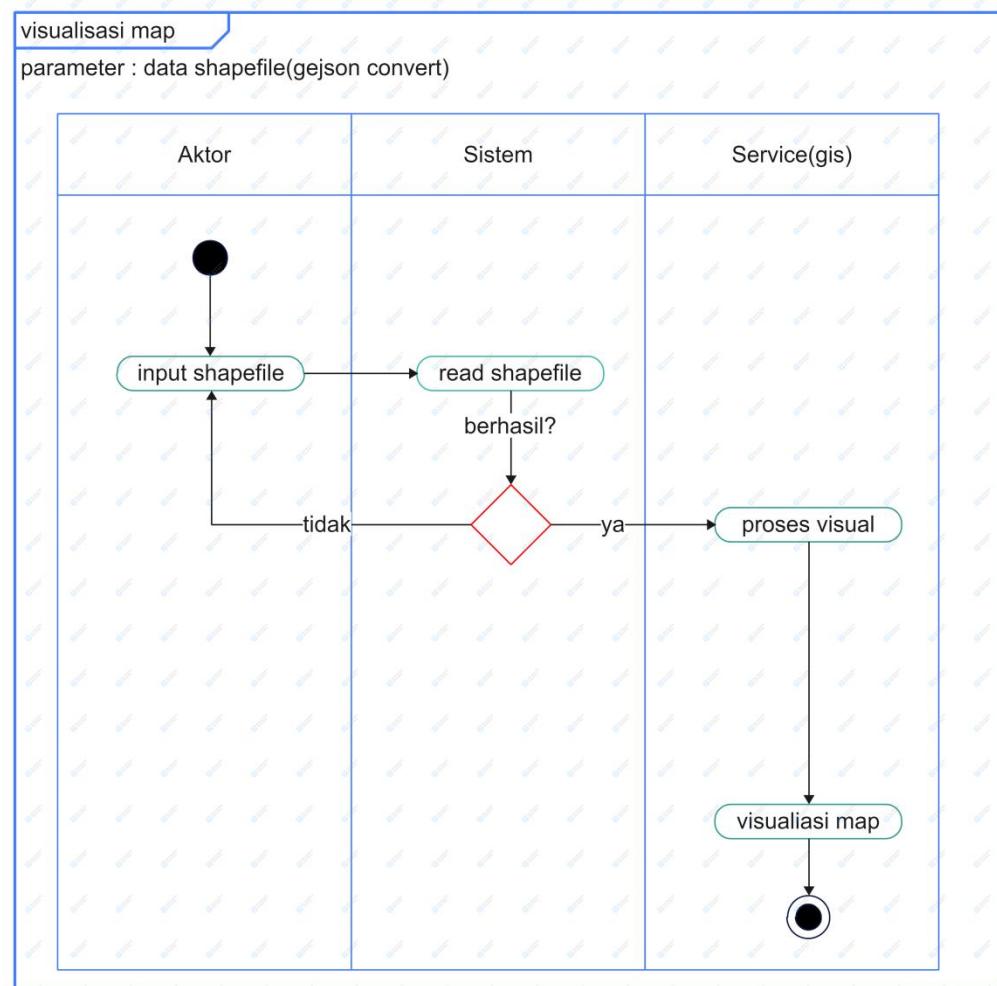
Pada Activity Diagram k-mean clustering menggambarkan urutan aktivitas antara sistem terhadap aksi yang dilakukan oleh user untuk melakukan clustering wilayah.



3.4 Activity Diagram K-mean Clustering

3.4.4. Activity Diagram visualiasi map

Pada Activity Diagram visualiasi map menggambarkan urutan aktivitas antara sistem terhadap aksi yang dilakukan oleh user untuk melakukan pemetaan wilayah yang sudah tercluster.



3.5 Activiy Diagram Visualiasi MAP

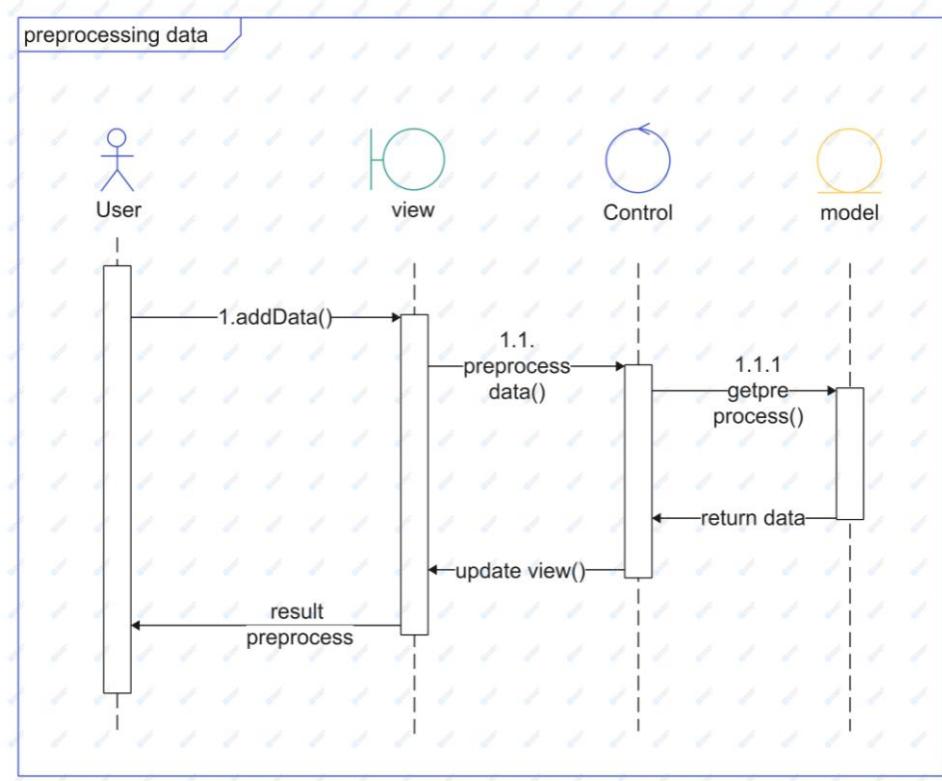
3.4.5. Sequence Diagram

Sequence diagram menggambarkan interaksi antar objek di dalam dan di sekitar sistem berupa pesan yang digambarkan terhadap waktu.

Sequence diagram terdiri atas dimensi vertikal (waktu) dan dimensi horizontal (objek-objek yang terkait).

3.4.5.1. Sequence Diagram Preprocessing

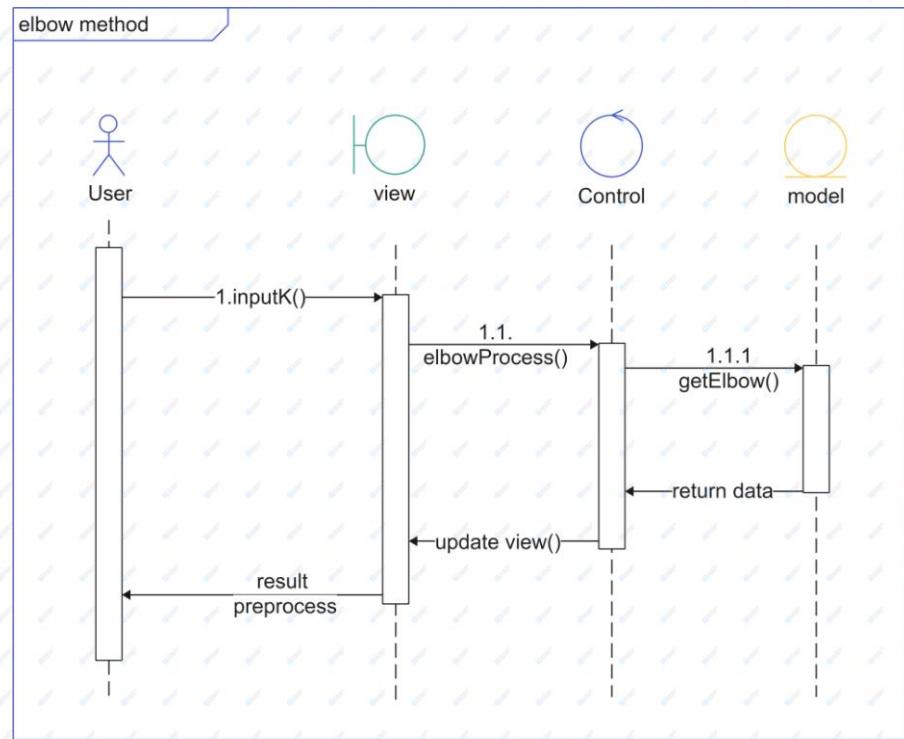
Pada sequence diagram preprocessing ini menggambarkan bagaimana sistem melakukan preprocessing , dimana terdapat tahapan dari user menambahkan data sampai dengan mendapat hasil dari preprocess.



3.6 Sequence Diagram Preprocessing

3.4.5.2. Sequence Diagram Elbow

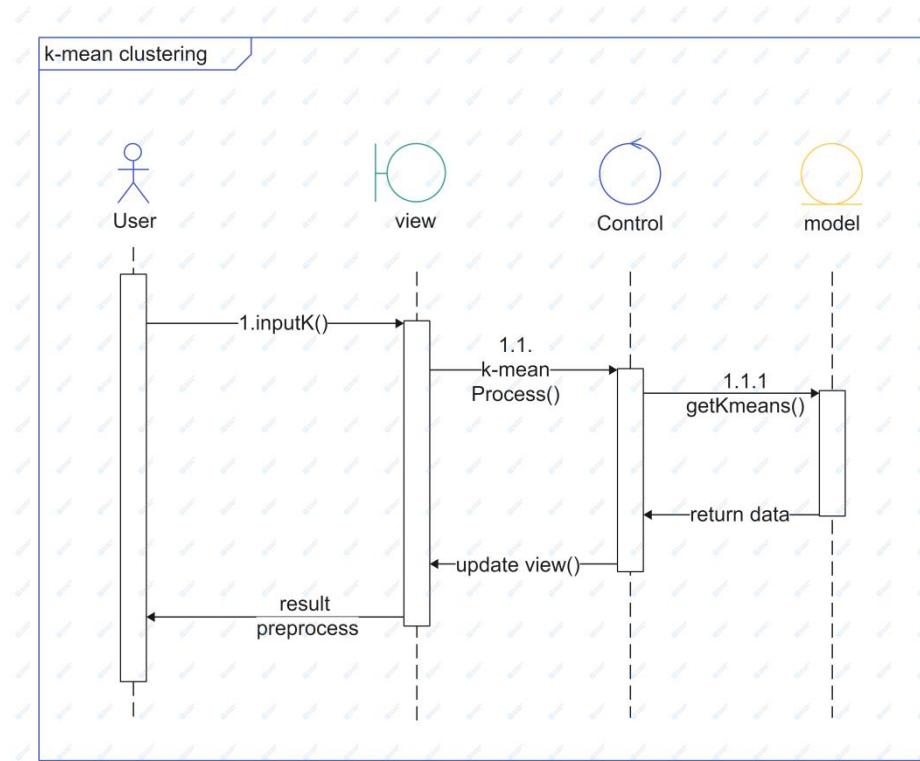
pada sequence diagram elbow ini menggambarkan bagaimana sistem melakukan penentuan jumlah k yang optimal, dimana terdapat tahapan sistem untuk melakukan elbow process sampai dengan menetukan jumlah k yang optimal(return data).



3.7 Sequence Diagram Elbow

3.4.5.3. Sequence Diagram K-mean clustering

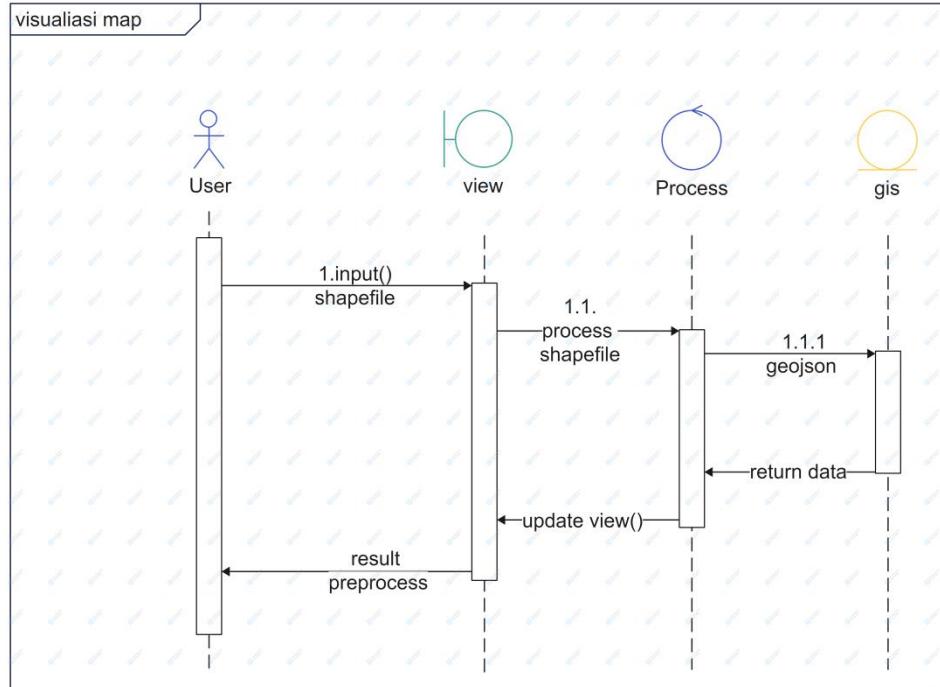
Sequence Diagram k-mean clustering menggambarkan bagaimana sistem melakukan clustering objek , dimana terdapat tahapan user melakukan input jumlah cluster hasilnya setiap wilayah akan ditentukan clusternya.



3.8 Sequence Diagram K-mean Clustering

3.4.5.4. Sequence Diagram Visualisasi Map

Sequence Diagram Visualisasi Map menggambarkan bagaimana sistem melakukan visualisasi map dari hasil clustering yang telah didapat, dimana terdapat tahapan untuk input data shapefile yang isinya berupa geojson(polygon wilayah) dan akan ditampilkan.

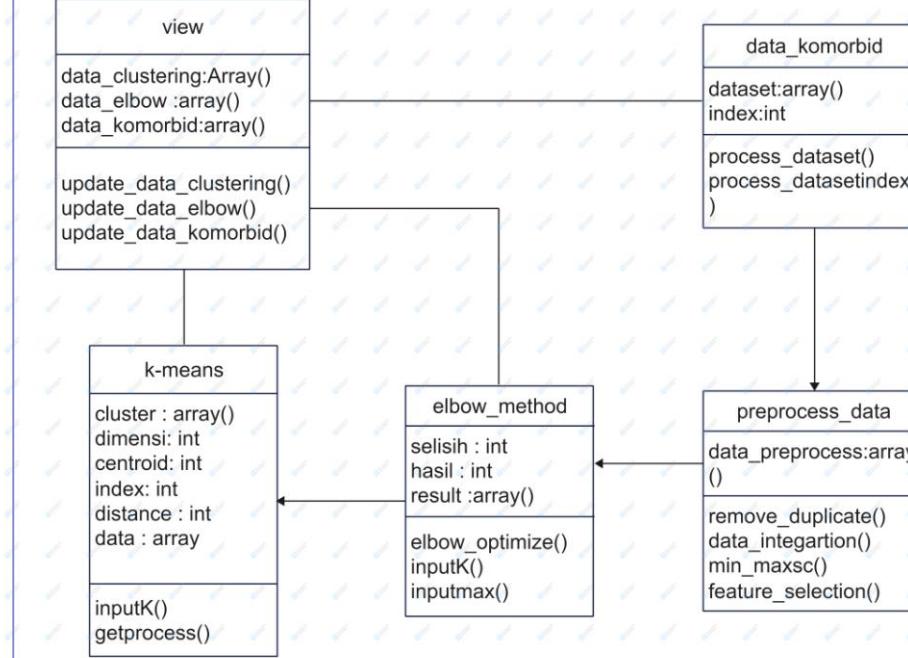


3.9 Sequence Diagram Visualiasi Map

3.4.6. Class Diagram

Class Diagram merupakan struktur statis sistem yang digambarkan kelas-kelas mewakili kelas objek dalam sistem dan hubungan antara masing-masing kelas, *package* serta objek yang berhubungan satu sama lain. Class Diagram ini berfungsi membantu dalam visualisasi struktur kelas dari suatu sistem.

class diagram sistem analisis kesehatan masyarakat yang rentan terhadap paparan covid-19



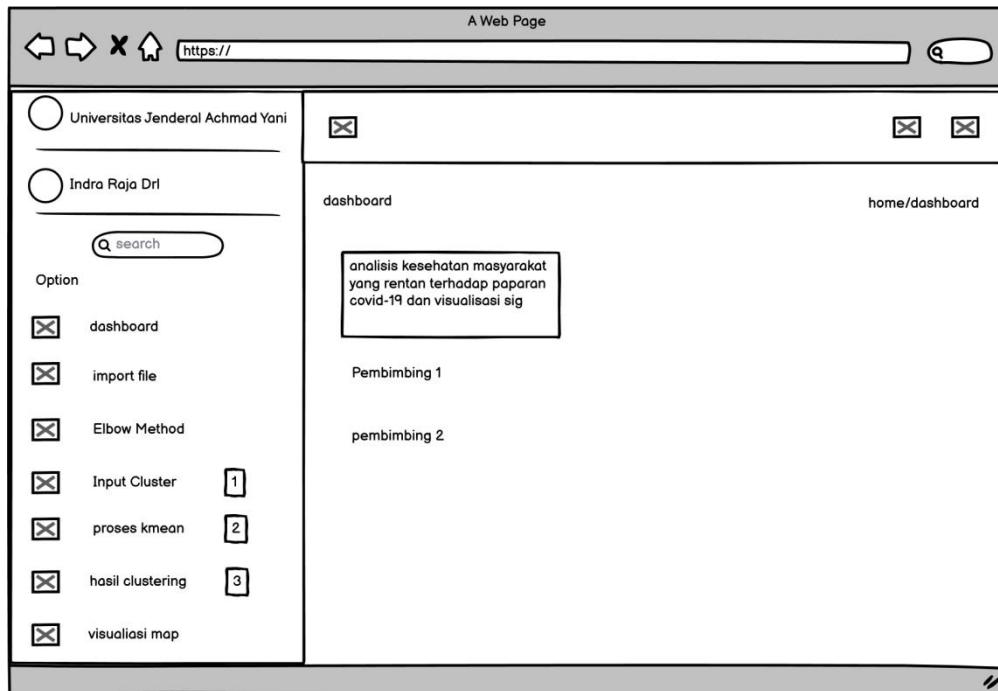
3.10 Class Diagram Sistem

3.8. Perancangan Antarmuka

Perancangan antarmuka adalah sebuah design sederhana mengenai tampilan yang akan nantinya diimplementasikan, pada antarmuka ini terdapat komponen-komponen , terdapat 7 halaman yaitu : dashboard,import file,elbow,kmean,input cluster,proses kmean dan visual map.

3.7.1. Perancangan Antarmuka Dashboard

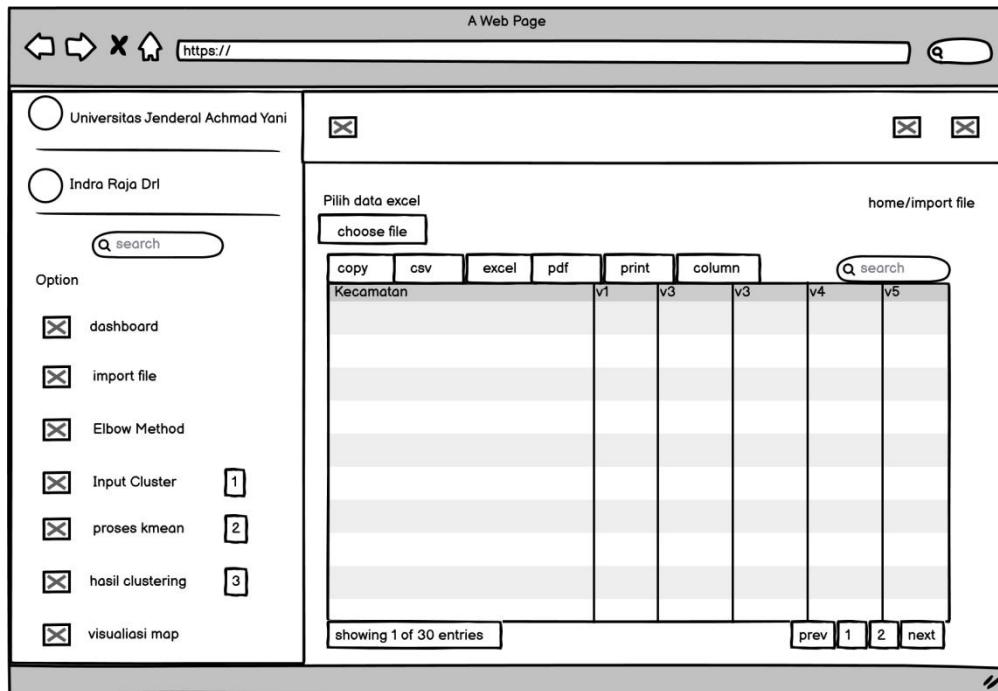
Perancangan antarmuka dashboard ini merupakan halaman awal dari user mengenai informasi judul tugas akhir dan identitas pembimbing 1 dan pembimbing2.



3.11 Gambar Perancangan Antarmuka Dashboard

3.7.2. Perancangan Antarmuka Import File

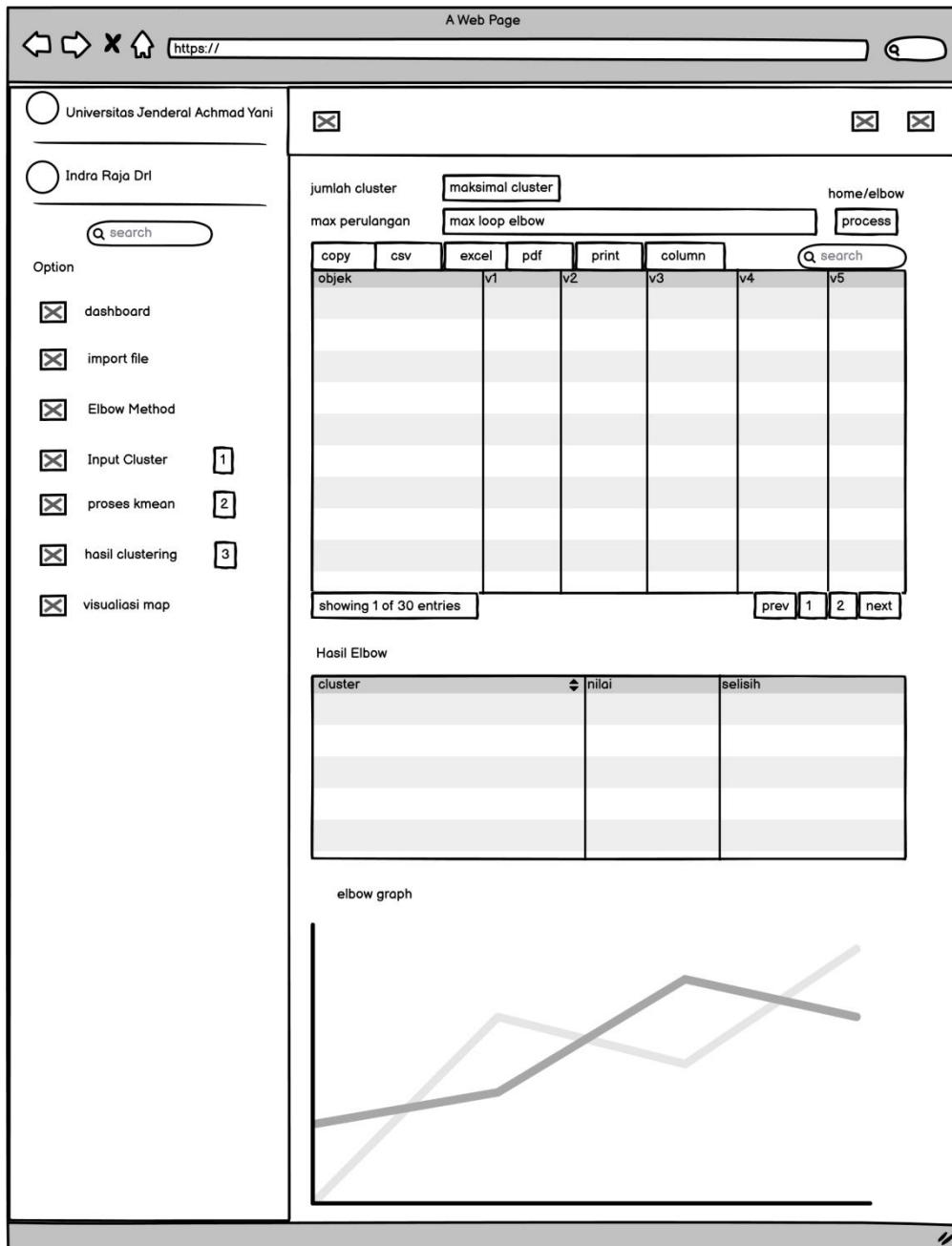
Perancangan antarmuka Import file ini merupakan halaman kedua dari sistem terdapat box import file dan table dari data yang akan tampilan, untuk fitur dari table terdiri dari tampilan search box, export dan paging.



3.12 Gambar Perancangan Antarmuka Import File

3.7.3. Perancangan Antarmuka Elbow

Perancangan antarmuka elbow ini merupakan halaman ketiga dari sistem untuk input jumlah cluster dan max perulangan, lalu terdapat perhitungan sse dan hasilnya serta menampilkan grafik elbow.



3.13 Gambar Perancangan Antarmuka Elbow

3.7.4. Perancangan Antarmuka Input jumlah cluster

Perancangan antarmuka input jumlah cluster ini merupakan halaman keempat ini dari sistem untuk menampilkan field type centroid, jumlah cluster dan max perulangan ,juga terdapat button process.

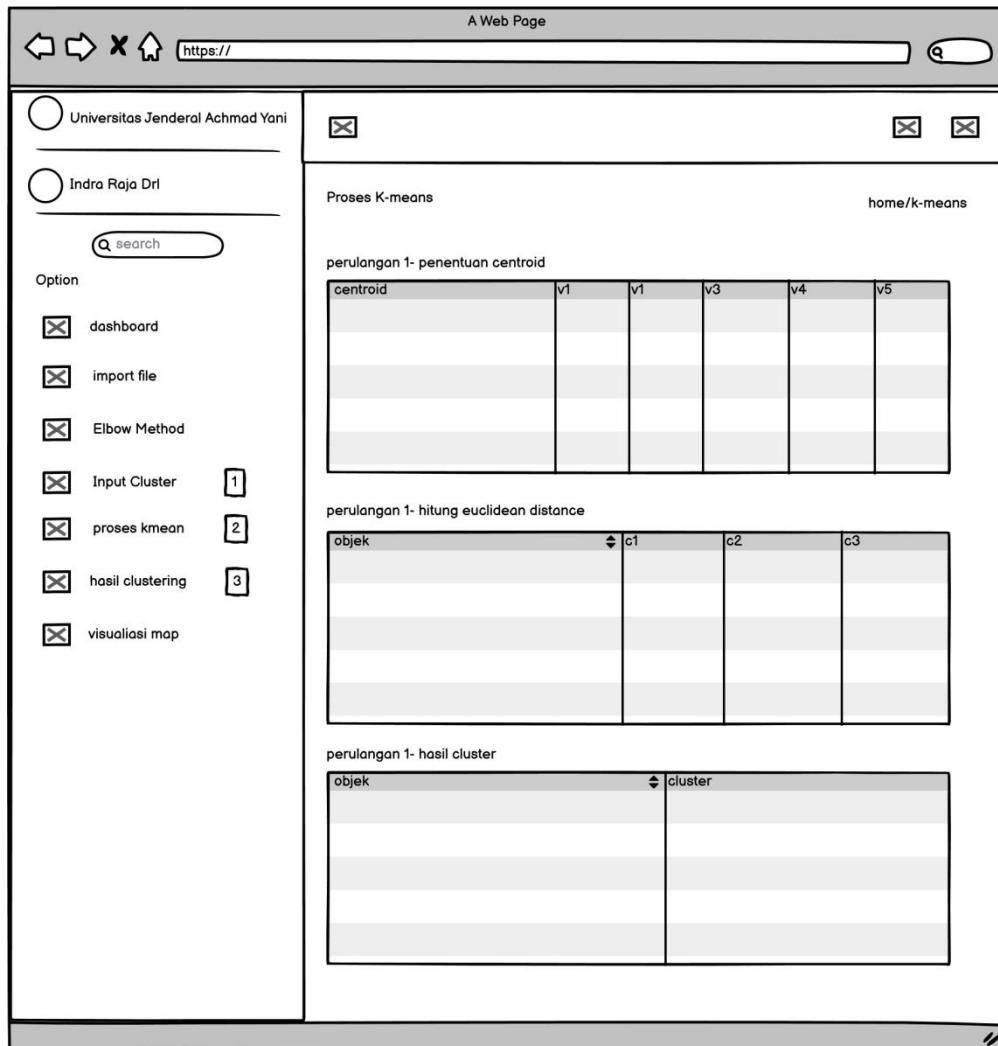
A Web Page

The screenshot shows a web application interface. On the left is a sidebar with user information ('Universitas Jenderal Achmad Yani' and 'Indra Raja DrI') and a search bar. Below this are several menu items: 'dashboard', 'import file', 'Elbow Method', 'Input Cluster' (with value '1'), 'proses kmean' (with value '2'), 'hasil clustering' (with value '3'), and 'visualiasi map'. The main panel is titled 'Input jumlah cluster' and contains three input fields: 'type centroid' (set to 'random centroid'), 'jumlah cluster' (set to 'jumlah cluster'), and 'max perulangan' (set to '10'). A 'process' button is located at the bottom right of this panel.

3.14 Gambar Perancangan Antarmuka Input Jumlah cluster

3.7.5. Perancangan Antarmuka Proses K-means

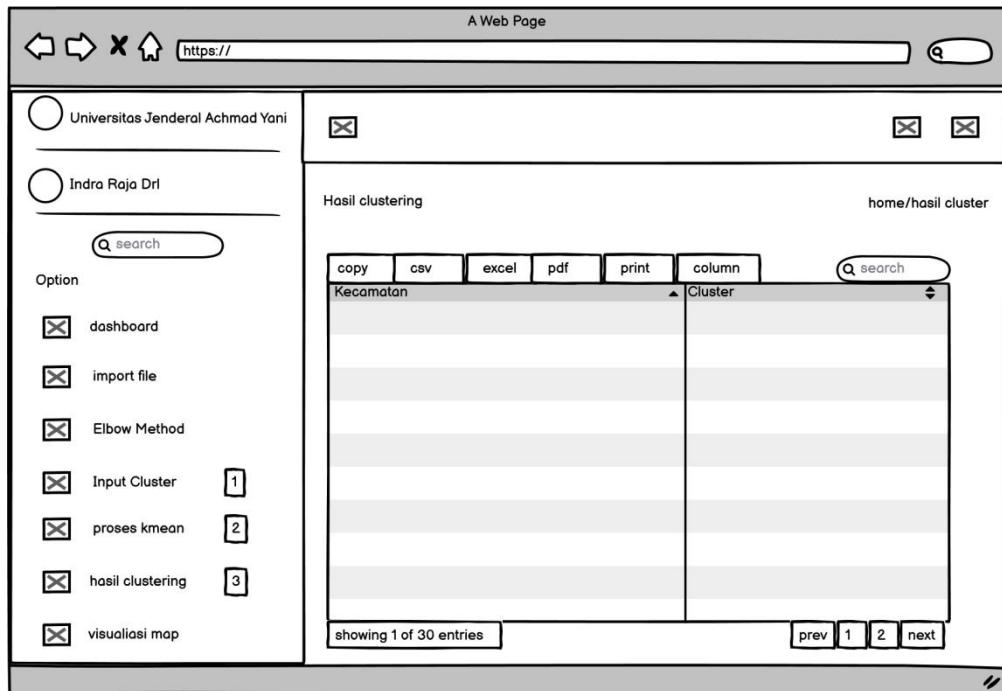
Perancangan antarmuka proses k-mean ini merupakan halaman kelima ini dari sistem terdapat tampilan table perulangan I- penetuan centroid,table perulangan I- euclidian distance dan hasil cluster.



3.15 Gambar Perancangan Antarmuka Proses K-Means

3.7.6. Perancangan Antarmuka Hasil Clustering

Perancangan antarmuka proses k-mean ini merupakan halaman keenam ini dari sistem terdapat tampilan table hasil cluster, dengan header kecamatan dan cluster.



3.16 Gambar Perancangan Antarmuka Hasil Clustering

BAB IV IMPLEMENTASI DAN PENGUJIAN

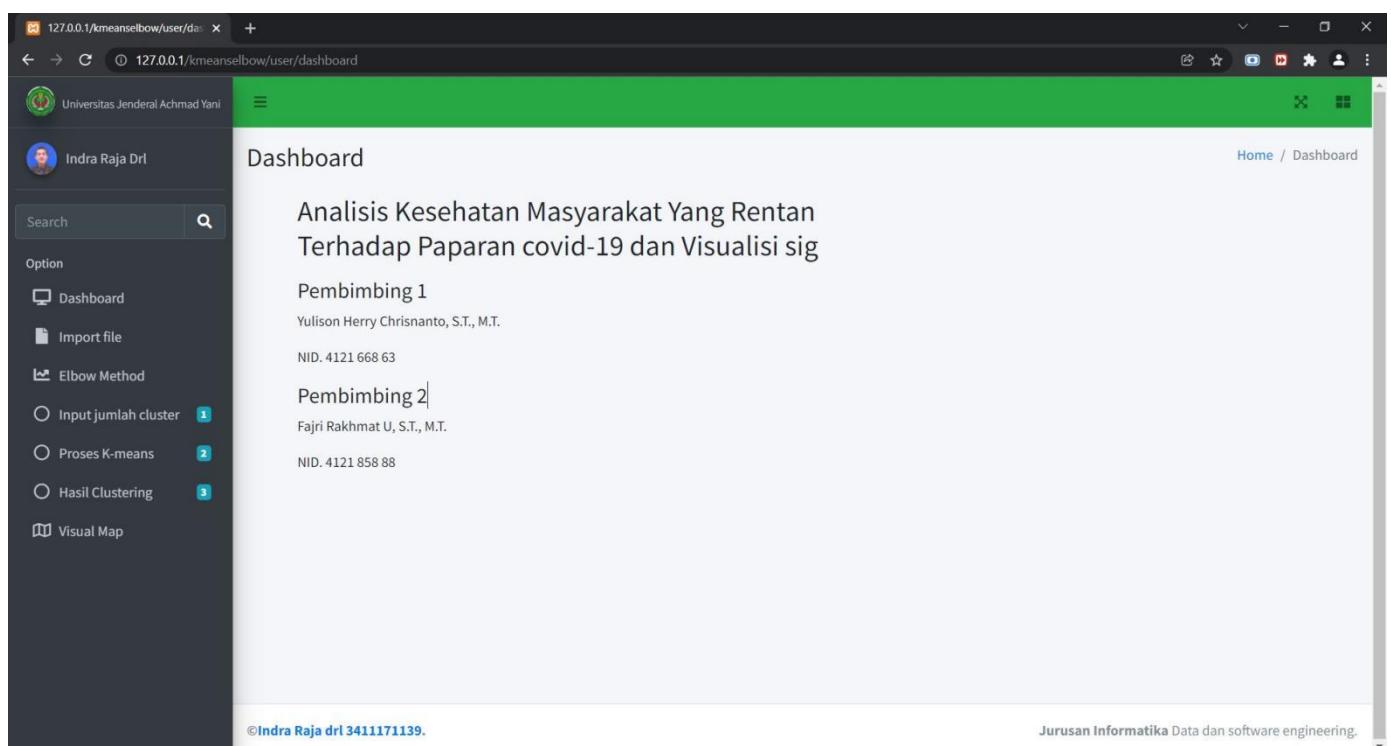
Pada bab IV ini berisi mengenai implementasi perangkat lunak berdasarkan sistem yang telah dirancang sebelumnya. Selain itu, dilakukan juga uji kualitas terhadap perangkat lunak.

4.1. Implementasi Sistem

Implementasi sistem menampilkan halaman dashboard,import file,elbow,input cluster,proses k-mean, hasil clustering dan visual map.

4.1.1. Implementasi Antarmuka dashboard

Implementasi antarmuka halaman dashboard ini merupakan tampilan yang menyajikan informasi judul tugas akhir ,nama dosen pembimbing dan nid pembimbing.



4.1 Gambar Implementasi Antarmuka

4.1.2. Implementasi Antarmuka import file

Implementasi antarmuka import file ini merupakan tampilan untuk melakukan import data dengan format xls. Data dengan format dimensi array akan ditampilkan dan disave untuk digunakan pada proses elbow dan kmeans.

The screenshot shows a web-based application for data analysis. On the left, there is a sidebar with a user profile (Indra Raja DrI) and navigation options: Dashboard, Import file, Elbow Method, Input jumlah cluster (with a count of 1), Proses K-means (with a count of 2), Hasil Clustering (with a count of 3), and Visual Map. The main content area has a title "Pilih data excel" and a sub-section "Choose File No file chosen". Below this is a table titled "DataTable komorbid covid-19". The table has the following columns: Kecamatan, Obesitas, Diabetes, Tuberculosis, Diare, Hipertensi, Jumlah_Penduduk, and positif_covid. The data rows are as follows:

Kecamatan	Obesitas	Diabetes	Tuberculosis	Diare	Hipertensi	Jumlah_Penduduk	positif_covid
Andir	0.128837209	0.859891259	1.905882353	0.53748232	0.917912476	1.298979914	1.637540453
Antapani	0.064883721	0.888331242	0.211764706	1.21301273	1.120956782	0.938602026	1.663430421
Arcamanik	0.041162791	1.519029695	0.423529412	0.849787836	1.293829845	0.905230133	1.760517799
Astanaanyar	0.449069767	0.332078628	0.705882353	0.017538897	0.583582495	0.855032075	0.919093851
Babakan Ciparay	0.358837209	0.273525721	2	0.772842999	1.047567274	2	0.886731392
Bandung Kidul	0	0.301129235	0.564705882	0.37397454	0.379450938	0.611052687	0.582524272
Bandung Kulon	0.083023256	0.60811376	1.341176471	0.540311174	0.472139168	1.893101272	2
Bandung Wetan	0.109302326	0.117105813	0.1411176471	0.036775106	0.172057624	0.065779788	0
Batununggal	0.410697674	2	0.752941176	0.418670438	1.902962762	1.683633049	0.705501618
Bojongloa Kaler	0.438604651	0.403178586	1.364705882	0.704950495	0.877684153	1.722981737	0.705501618

Showing 1 to 10 of 30 entries

4.2 Gambar Implementasi Antarmuka Import File

4.1.3. Implementasi Antarmuka elbow

1. Table Perhitungan Elbow

Implementasi Antarmuka elbow ini merupakan halaman untuk melakukan input jumlah cluster yang akan dihitung dan max perulangan sehingga dapat dilakukan proses perhitungan sum of square error.

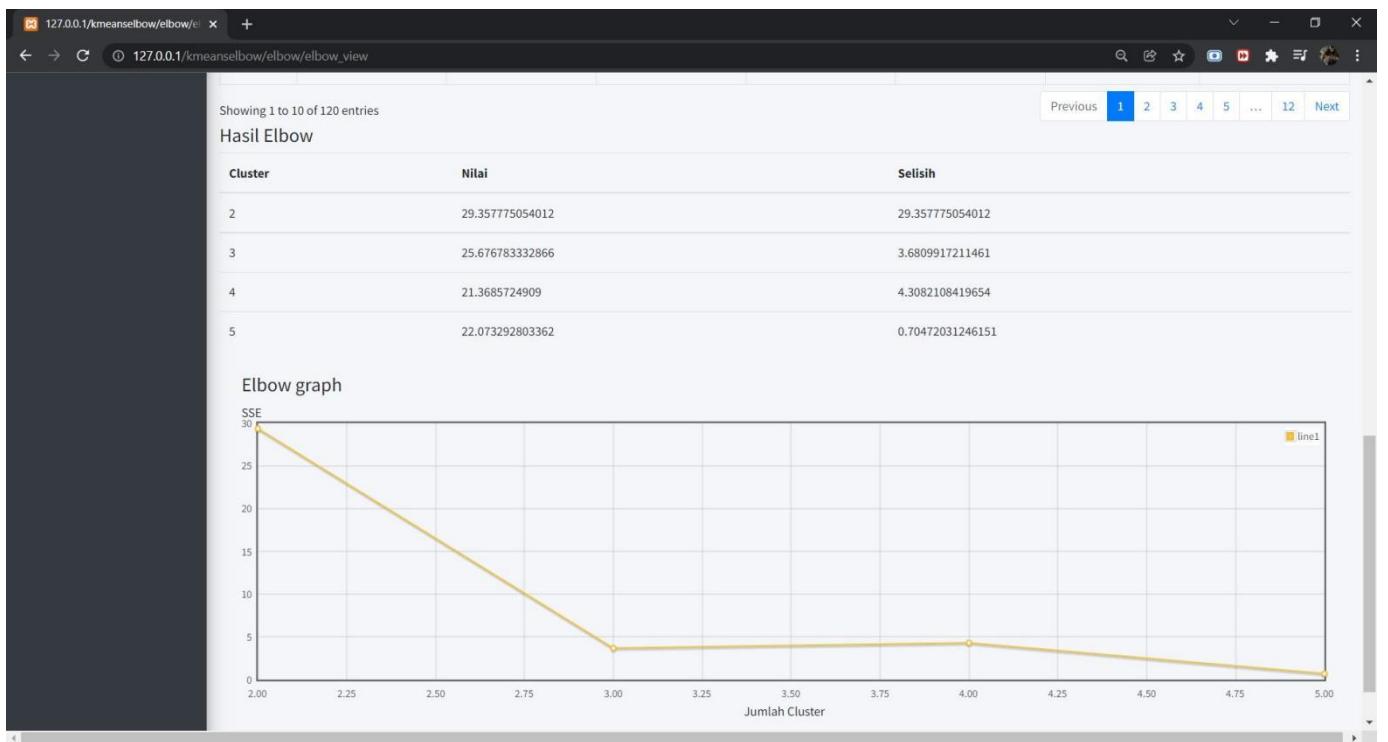
The screenshot shows a web-based application for the Elbow method. On the left, there's a sidebar with a logo, user name 'Indra Raja Dr.', and a search bar. Below the search bar are several menu items: Dashboard, Import file, Elbow Method, Input jumlah cluster (with a count of 1), Proses K-means (with a count of 2), Hasil Clustering (with a count of 3), and Visual Map. The main content area has a green header 'Elbow method'. It contains two input fields: 'Jumlah Cluster (Dari - Sampai)' and 'Maksimal Cluster' (set to 2), and 'Max Perulangan' with a dropdown set to 'max loop elbow'. A blue 'Process' button is next to it. Below these is a table titled 'Table Perhitungan Elbow' with columns: Cluster, Obesitas, Diabetes, Tuberculosis, Diare, Hipertensi, Jumlah_Penduduk, and positif_covid. The table rows represent different cluster configurations from 2 to 6, showing complex mathematical formulas for each metric.

Cluster	Obesitas	Diabetes	Tuberculosis	Diare	Hipertensi	Jumlah_Penduduk	positif_covid
Cluster - 2	(0.128837209 - 0.2575872091875)^2	(0.859891259 - 0.7871183605)^2	(1.905882353 - 0.9926470588125)^2	(0.53748232 - 0.8419377651875)^2	(0.917912476 - 1.13293354175)^2	(1.298979914 - 1.3911141199375)^2	(1.637540453 - 1.2394822006875)^2
Cluster - 2	(0.064883721 - 0.2575872091875)^2	(0.888331242 - 0.7871183605)^2	(0.211764706 - 0.9926470588125)^2	(1.21301273 - 0.8419377651875)^2	(1.120956782 - 1.13293354175)^2	(0.938602026 - 1.3911141199375)^2	(1.663430421 - 1.2394822006875)^2
Cluster - 2	(0.041162791 - 0.2575872091875)^2	(1.519029695 - 0.7871183605)^2	(0.423529412 - 0.9926470588125)^2	(0.849787836 - 0.8419377651875)^2	(1.293829845 - 1.13293354175)^2	(0.905230133 - 1.3911141199375)^2	(1.760517799 - 1.2394822006875)^2
Cluster - 2	(0.449069767 - 0.12524916942857)^2	(0.332078628 - 0.29156957635714)^2	(0.705882353 - 0.36134453771429)^2	(0.017538897 - 0.39050313185714)^2	(0.583582495 - 0.46843086228571)^2	(0.855032075 - 0.57903540014286)^2	(0.919093851 - 0.58298659264286)^2
Cluster - 2	(0.358837209 - 0.2575872091875)^2	(0.273525721 - 0.7871183605)^2	(2 - 0.9926470588125)^2	(0.772842999 - 0.8419377651875)^2	(1.047567274 - 1.13293354175)^2	(2 - 1.3911141199375)^2	(0.886731392 - 1.2394822006875)^2
Cluster - 2	(0 - 0.12524916942857)^2	(0.301129235 - 0.29156957635714)^2	(0.564705882 - 0.36134453771429)^2	(0.37397454 - 0.39050313185714)^2	(0.379450938 - 0.46843086228571)^2	(0.611052687 - 0.57903540014286)^2	(0.582524272 - 0.58298659264286)^2

4.3 Gambar Implementasi Antarmuka Elbow 1

2. Hasil Elbow dan Grafik elbow

Implementasi Antarmuka hasil elbow ini merupakan halaman untuk menampilkan hasil perhitungan jarak sehingga dapat diketahui nilai sse dan selisihnya, dari perbandingan cluster-I dengan cluster-I apabila memiliki selisih sse yang drastis , maka akan membentuk grafik elbow sehingga dapat ditentukan jumlah cluster yang optimalnya.



4.4 Gambar Implementasi Antarmuka Elbow 1

4.1.4. Implementasi Antarmuka input cluster

Implementasi antarmuka input cluster ini merupakan halaman untuk input jumlah cluster yang optimal didapat dari hasil elbow juga input max perulangan untuk penentuan centroid, terdapat button simpan untuk menyimpan data cluster ke session.

Universitas Jenderal Achmad Yani

Indra Raja DrI

Search

Option

Dashboard

Import file

Elbow Method

Input jumlah cluster 1

Proses K-means 2

Hasil Clustering 3

Visual Map

Type Centroid: Random Centroid

Jumlah Cluster: 3

Max Perulangan: 10

Simpan

©Indra Raja drI.

Jurusan Informatika Data dan software engineering.

4.5 Gambar Implementasi Antarmuka Input Cluster

4.1.5. Implementasi Antarmuka Proses K-means

1. Implementasi Antarmuka Penetuan Centroid & Euclidean distance

Implementasi antarmuka penentuan centroid & proses ED ini merupakan halaman penentuan centroid secara random dari jumlah cluster yang telah ditentukan, setelah penentuan centroid akan dihitung jarak cluster dengan objek

The screenshot shows a web-based K-means clustering application. On the left, a sidebar lists navigation options: Dashboard, Import file, Elbow Method, Input jumlah cluster (selected), Proses K-means (selected), Hasil Clustering (selected), and Visual Map. The main content area has a green header bar with the title 'Proses K-means'. Below it, there are two tables.

Perulangan 1 - Penentuan Centroid

Centroid	Obesitas	Diabetes	Tuberculosis	Diare	Hipertensi	Jumlah_Penduduk	positif_covid
Centroid 1	0.410697674	2	0.752941176	0.418670438	1.902962762	1.683633049	0.705501618
Centroid 2	0.226744186	0.441656211	0.211764706	1.162659123	0.40418592	0.782609458	0.724919094
Centroid 3	0.12255814	0.110414053	0.164705882	0.273833098	1.168524055	0.208942405	0.459546926

Perulangan 1 - Hitung Euclidean Distance

Andir	2.1701754247349	2.1929017614748	2.5103024934569
Antapani	2.0888940939524	1.2834136277824	1.8646060396611
Arcamanik	1.6589984742451	1.793564504599	2.1406961632136
Astanaanyar	2.3279863985349	1.3007768811618	1.2186367724549
Babakan Ciparay	2.3513040162684	2.3059967527819	2.6654742570536
Bandung Kidul	2.5649376295051	0.93170351098917	1.0103828864459
Bandung Kulon	2.4844303602327	2.1388998937783	2.7200551999002
Bandung Wetan	3.2042852678882	1.5763752993563	1.1320754750554

4.6 Gambar Implementasi Antarmuka Perulangan-I 1

2. Implementasi Antarmuka Perulangan-i Hasil Clustering

Implementasi antarmuka perulangan-I hasil clustering ini merupakan halaman untuk menampilkan jarak terdekat antara objek dengan clusternya lalu didapat clusternya dan akan dihitung kembali centroid baru sampai centroidnya tidak berubah.

Perulangan 1 - Hasil Cluster	
Andir	1
Antapani	2
Arcamanik	1
Astanaanyar	3
Babakan Ciparay	2
Bandung Kidul	2
Bandung Kulon	2
Bandung Wetan	3
Batununggal	1
Bojongloa Kaler	2
Bojongloa Kidul	2
Buahbatu	3
Cibeunying Kaler	2
Cibeunying Kidul	2
Cibiru	2

4.7 Gambar Implementasi Antarmuka Perulangan -I 1

4.1.6. Implementasi Antarmuka Hasil clustering

Implementasi antarmuka hasil clustering ini merupakan halaman untuk menampilkan cluster tetap dari perhitungan k-mean.

Hasil Clustering	
Copy	CSV
Excel	PDF
Print	Column visibility ▾
Search:	Search: <input type="text"/>
Kecamatan	Cluster
Andir	1
Antapani	1
Arcamanik	1
Astanaanyar	2
Babakan Ciparay	3
Bandung Kidul	2
Bandung Kulon	1
Bandung Wetan	2
Batununggal	1
Bojongloa Kaler	3

4.8 Gambar Implementasi Antarmuka Hasil clustering

4.2. Pengujian Perangkat Lunak

Pengujian perangkat lunak dilakukan untuk uji kesesuaian antara rancangan dan hasil yang diharapkan. Tujuan dalam pengujian kualitas ini adalah untuk menentukan kesesuaian unit-unit, merancang pengujian kualitas, pelaksanaan pengujian dan kesimpulan dari hasil pengujian.

4.2.1. Metode Pengujian

Metode pengujian yang digunakan yaitu *Black Box Testing*, pengujian *black box* berfokus pada persyaratan fungsional perangkat lunak yang telah dibangun dan menganalisis sistem dalam memperoleh kumpulan masukan yang dikerjakan oleh fungsional program. Tujuan metode *black box* yaitu mencari penyebab apabila terjadi kesalahan pada:

1. Fungsi yang salah atau hilang,
2. Kesalahan pada antarmuka,
3. Kesalahan performa,
4. Kesalahan inisialisasi dan tujuan akhir.

4.2.2. Tahap Pengujian Black Box

Tahapan Pengujian perangkat lunak menggunakan *black box testing* berdasarkan perancangan sebelumnya seperti mengelompokkan proses-proses yang mengacu pada perancangan sebelumnya. Pada tahapan ini, pengelompokan proses-proses disesuaikan pada perancangan pada BAB III, diantaranya:

1. *Use Case Preprocess Data*
2. *Use Case Import file*
3. *Use Case Menampilkan Data Set*
4. *Use Case Search data*

5. Use Case Input jumlah K elbow
6. Use Case Process Sum of Square Error
7. Use Case Menampilkan grafik Elbow
8. Use Case Menampilkan Selisih dan SSE
9. Use case Input Jumlah cluster K-Means
10. Use case Penetuan centroid
11. Use case Proses euclidean distance
12. Use case Penentuan Cluster

4.2.3. Tujuan Pengujian

Tahap ini membahas mengenai tujuan pengujian kualitas terhadap perangkat lunak yang dibangun, tujuan tersebut dapat dilihat pada tabel Tujuan Pengujian pada tabel 4.1.

4.1 Tabel Tujuan Pengujian

No	Use Case	Tujuan Pengujian
1	<i>Use Case Preprocessing Data</i>	Menguji kemampuan sistem untuk melakukan preprocessing data
2	Use Case Import File	Menguji kemampuan sistem untuk melakukan import data dan disimpan disistem
3	<i>Use Case Menampilkan Data Set</i>	Menguji kemampuan sistem untuk menampilkan data set yang disimpan dalam session.
4	<i>Use Case Search Data</i>	Menguji kemampuan sistem untuk melakukan search data sehingga memudahkan untuk menampilkan data yang dicari

5	<i>Use Case Input Jumlah k elbow</i>	Menguji kemampuan sistem untuk melakukan input jumlah k elbow
5	<i>User Case Process Sum of Square error</i>	Menguji Kemampuan sistem untuk melakukan process sum of square error
6	<i>Use Case Menampilkan Grafik Elbow</i>	Menguji Kemampuan sistem untuk menampilkan grafik elbow
7	<i>Use Case Menampilkan Selisih dan SSE</i>	Menguji Kemampuan sistem untuk menampilkan data nilai selisih dan sse perhitungan sum of square error
8	<i>Use case Input Jumlah cluster K-Means</i>	Menguji Kemampuan sistem untuk melakukan input data numerik sebagai penentuan jumlah cluster
9	<i>Use case Penetuan centroid</i>	Menguji Kemampuan sistem untuk menentukan centroid random diawal dan menentukan centroid baru dengan rata-rata.
10	<i>Use case Proses euclidean distance</i>	Menguji Kemampuan sistem menghitung euclidean distance sehingga dari perhitungan dapat menentukan jarak terdekat untuk objek.
11	<i>Use case Penentuan Cluster</i>	Menguji Kemampuan sistem untuk menentukan cluster wilayah sampai dengan hasil cluster tetap didapatkan

4.2.4. Kategori Hasil Pengujian

Tahap dalam menentukan kategori keberhasilan terhadap pengujian sistem, dibagi ke dalam dua kategori yaitu sesuai dan tidak sesuai, dengan penjelasan sebagai berikut:

1. Sesuai

Hasil pengujian termasuk dalam kategori sesuai jika sistem yang diuji kualitasnya memenuhi tujuan dari penggunaan yang ditetapkan, serta sesuai dengan perancangan sistem dan kegunaannya.

2. Tidak Sesuai

Hasil pengujian termasuk dalam kategori tidak sesuai, jika sistem yang diuji kualitasnya tidak memenuhi tujuan dari perancangan sistem dan kegunaannya.

4.2.5. Perancangan Pengujian Perangkat Lunak

Perancangan pengujian kualitas digunakan sebagai acuan untuk mengetahui kualitas dari sistem yang telah dibangun. Penjelasan perancangan pengujian dapat diliha pada Tabel 4.2.

4.2 Table Pengujian Perangkat Lunak

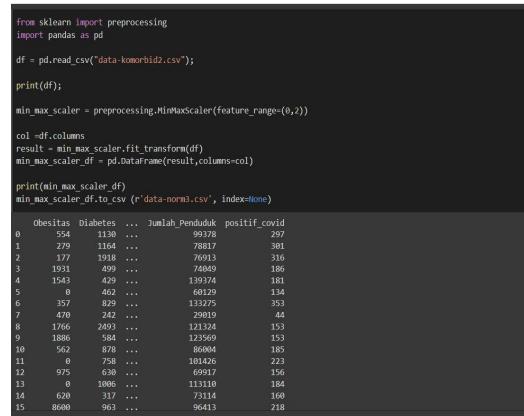
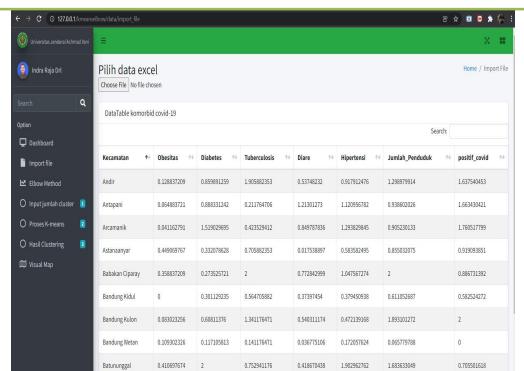
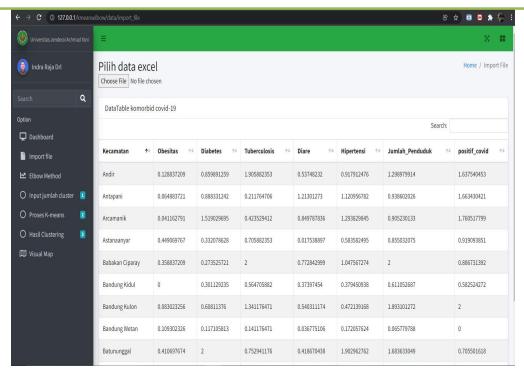
No	Use Case	Kode Uji	Deksripsi
1	<i>Use Case Preprocessing Data</i>	K.U.1	Uji preprocess sehingga data memiliki perubahan
2	Use Case Import File	K.U.2	Uji input data sehingga data dapat terbaca
3	<i>Use Case Menampilkan Data Set</i>	K.U.3	Uji menampilkan data pada sistem sesuai dengan data set yang dimiliki pada excel.
4	<i>Use Case Search Data</i>	K.U.4	Uji mencari data sesuai dengan keyword yang dicari
5	<i>Use Case Input Jumlah kelbow</i>	K.U.5	Uji type inputan sehingga hanya type numerik yang dapat diinput
5	<i>User Case Process Sum of Square error</i>	K.U.6	Uji pemrosesan apakah sudah sesuai kalkulasi perhitungan data dengan centroid.

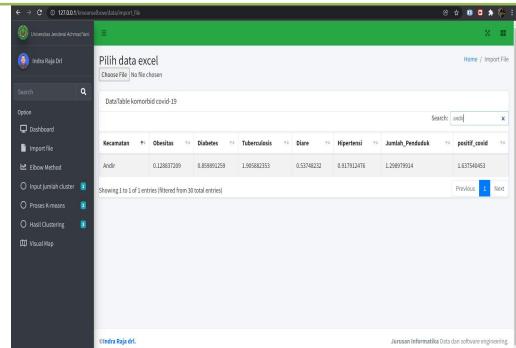
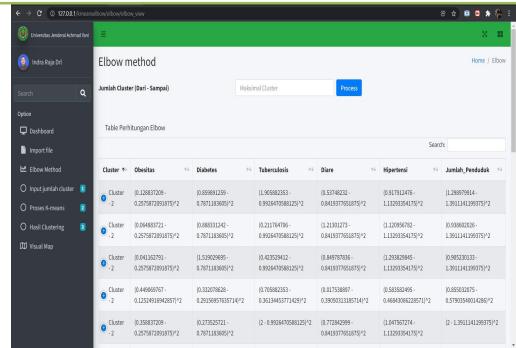
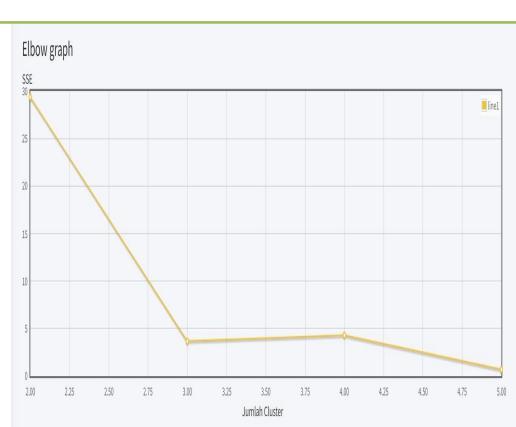
6	<i>Use Case Menampilkan Grafik Elbow</i>	K.U.7	Uji tampilan grafik chart line yang disesuaikan dengan hasil SSE dan jumlah cluster
7	Use Case Menampilkan Selisih dan SSE	K.U.8	Uji data SSE dan selisih ditampilkan hasilnya.
8	Use case Input Jumlah cluster K-Means	K.U.9	Uji type inputan sehingga hanya type numerik yang dapat diinput
9	Use case Penetuan centroid	K.U.10	Uji Data centroid yang didapat dari data hasil import.
10	Use case Proses euclidean distance	K.U.11	Uji proses perhitungan euclidian distance data set dan centroid.
11	Use case Penentuan Cluster	K.U.12	Uji penentuan cluster diambil dari penentuan centroid terdekat.

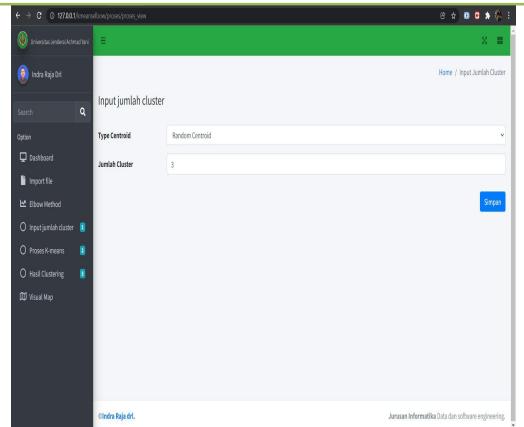
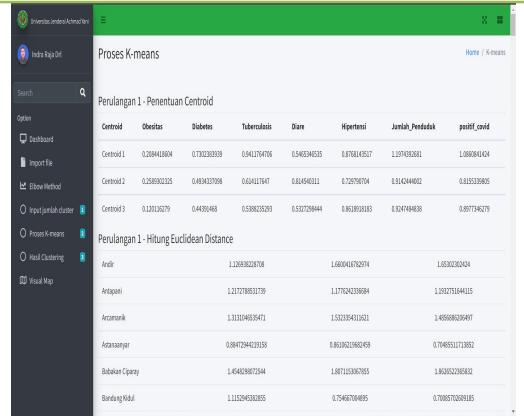
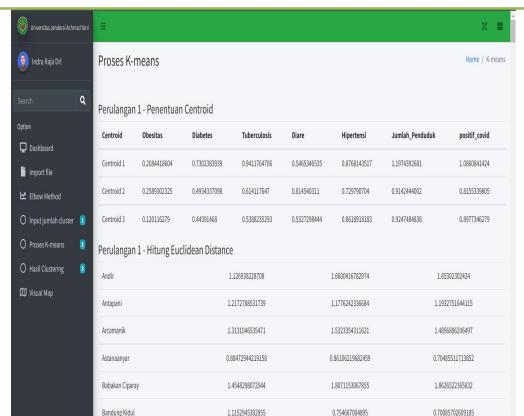
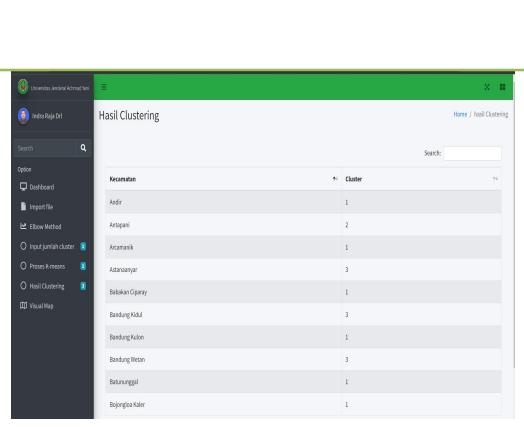
4.2.6. Pengujian Fungsionalitas Perangkat Lunak

Pengujian dilakukan dengan membandingkan kesesuaian perangkat lunak yang dibangun terhadap kebutuhan perangkat lunak yang dirancang hingga menghasilkan pengujian yang sesuai atau tidak sesuai. Pengujian sistem ditunjukkan pada Tabel 4.3.

4.3 Table Pengujian Fungsional Perangkat Lunak

Kod e Uji	Reaksi Sistem	Hasil Yang diharapkan	Hasil Uji	Ket
K.U. 1	Sistem mengambil data set yang telah diimport dan melakukan preprocessing	Sistem dapat melakukan preprocesssing dari hasilnya data diexport untuk digunakan	<pre>from sklearn import preprocessing import pandas as pd df = pd.read_csv("data-komorbid2.csv"); print(df); min_max_scaler = preprocessing.MinMaxScaler(feature_range=(0,1)) col=df.columns result = min_max_scaler.fit_transform(df) min_max_scaler_df = pd.DataFrame(result,columns=col) print(min_max_scaler_df) min_max_scaler_df.to_csv ("r'"+data-norm.csv'", index=None)</pre> 	sesuai
K.U. 2	Sistem menerima data yang sudah dilakukan preprocessing	Import data berhasil, sistem menampilkan seluruh data yang diimport		sesuai
K.U. 3	Sistem menerima data yang sudah disimpan	Sistem berhasil menampilkan data		sesuai

K.U. 4	Sistem menerima inputan yaitu kata kunci yang dicari	Sistem hanya menampilkan data yang dicari		sesuai
K.U. 5	Sistem menerima inputan user untuk menentukan jumlah cluster pada elbow	Sistem dapat menerima apa yang diperlukan oleh user		sesuai
K.U. 6	Sistem mengambil data SSE dan data cluster untuk ditampilkan angkanya	Sistem dapat menampilkan jumlah SSE dan jumlah cluster sehingga dapat dilihat grafik chart line		sesuai
K.U. 7	Sistem menerima hasil perhitungan sum of square error dan didapat penjumlahannya	Sistem dapat menampilkan hasil SSE dan selisih untuk tiap cluster		sesuai

K.U. 8	Sistem dapat menerima inputan dari user untuk menentukan jumlah cluster bertipe numerik	Sistem hanya dapat menenerima inputan bertipe numerik		sesuai
K.U. 9	Sistem menetukan centroid dari jumlah cluster yang ditentukan user	Sistem sudah sesuai menentukan centroid sesuai dengan data set yang dimiliki		sesuai
K.U. 10	Sistem menerima data set untuk dilakukan perhitungan menggunakan centroid	Sistem dapat melakukan perhitungan sesuai dengan data set dan centroid untuk tiap variable		sesuai
K.U. 11	Sistem menerima penentuan centroid yang tidak memiliki perubahan dengan	Sistem menetukan hasil clustering dan menampilkan cluster tetap.		sesuai

	centroid sebelumnya			
--	------------------------	--	--	--

4.2.7. Analisis Pengujian Black Box Testing

Berdasarkan hasil pengujian *black box testing* yang telah dilakukan pada Tabel dengan melakukan percobaan pada setiap fungsi yang ada, maka dapat dihitung persentase kesesuaian sistem dengan fungsi kesesuaian sebagai berikut:

Jumlah Kode Uji = 11 Kode Uji

Kode Uji dengan hasil Sesuai = 11 Kode Uji

Kode Uji dengan hasil Tidak Sesuai = 0 Kode Uji

$$\text{Persentase pengujian} = \frac{\text{jumlah kode uji sesuai} - \text{kode uji tidak sesuai}}{\text{jumlah kode uji}} \times 100\%$$

$$= \frac{11-0}{11} \times 100\%$$

$$= \frac{11}{11} \times 100\%$$

$$= 1 \times 100\%$$

$$= 100\%$$

Persentase = 100 % dihasilkan dari kode uji hasil yang sesuai / jumlah kode uji. Dari hasil perhitungan fungsi kesesuaian sistem dapat disimpulkan bahwa pengujian terhadap sistem clustering wilayah yang rentah terhadap paparan covid-19 menggunakan *Black Box Testing* telah berjalan sesuai dengan spesifikasi yang telah ditetapkan dengan persentase 100 % memenuhi spesifikasi.

4.3. Pengujian Cluster

Pengujian dilakukan terhadap semua data mahasiswa sebanyak 100 data dengan memasukkan beberapa titik pusat secara acak kemudian

dilakukan perhitungan nilai silhouette coefficient dari setiap titik pusat yang dimasukkan. Berikut adalah tabel proses pengujian terhadap data mahasiswa dengan memasukkan sebanyak tujuh cluster yang berbeda dapat dilihat pada Tabel 4.4.

4.4 Table Pengujian cluster

Jumlah Cluster	Hasil Silhouette Coefficient
2	0.2949
3	0.3219
4	0.2206
5	0.2248
6	0.2281
7	0.2050
8	0.1980

Dari hasil pengujian delapan cluster yang dimasukkan didapatkan cluster dengan jumlah 3 (tiga) yang memiliki nilai silhouette coefficient paling mendekati dengan nilai $S_i = 1$ yaitu dengan nilai 0.3219. Nilai silhouette dapat dikatakan baik apabila bernilai positif karena titik pusat sudah berada di dalam cluster yang tepat sedangkan jika nilai silhouette dikatakan negatif menandakan terjadinya overlapping sehingga titik pusat berada diantara dua cluster.

BAB V KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian tugas akhir ini , dapat disimpulkan bahwa wilayah dibagi berdasarkan karakteristik clusternya , setiap cluster memiliki karakteristik yang berbeda, untuk wilayah yang termasuk kepada cluster 1 yaitu andir ,arcamanik,babakan ciparay , bandung kulon,batununggal ,bojongloa kaler , bojongloa kidul , cibeunying kidul, coblong , kiara condong dan sukajadi memiliki karakteristik nilai rata-rata variable diabetes , tuberculosis,diare,hipertensi , jumlah penduduk dan angka positif covid yang lebih tinggi dari cluster 2 dan 3. untuk wilayah yang termasuk kepada cluster 2 antapani,buah batu, cibeunying kaler ,cibiru , cicendo,regol,sukasari dan ujung berung memiliki karakteristik rata-rata variable,obesitas yang lebih tinggi dari cluster 1 dan 3. untuk wilayah yang termasuk kepada cluster 3 yaitu Astaanyar , bandung kidul , bandung wetan, cidadap, cinambo,gedebage,lengkong,mandalajati,panyileukan, rancasari dan sumur bandung memiliki karakteristik nilai rata-rata variable obesitas ,diabetes,tuberculosis,diare,hipertensi,jumlah penduduk dan positif covid lebih kecil dari cluster 2 dan 1. Penentuan jumlah cluster yang optimal menggunakan data set yang dimiliki berjumlah 30 data dan 7 variable dengan ditentukan cluster maksimal sebanyak 5, hasilnya didapat bahwa jumlah k yang optimal adalah sebesar 3 karena terjadinya penurunan nilai SSE yang signifikan dari cluster 2 ke cluster 3.

5.2. Saran

Berdasarkan hasil dari penelitian ini terdapat beberapa saran untuk dijadikan referensi seperti :

1. Menggunakan jumlah variasi variable relasi covid yang banyak sehingga dapat menganalisis pengaruh jumlah kasus covid-19 yang tinggi terhadap akurasi penentuan variable yang berpengaruh terhadap covid-19

2. Perlu menggunakan data yang lebih banyak seperti mencangkup wilayah perkelurahan pada suatu kota.

REFERENSI

- [1] Bag, R., Ghosh, M., Biswas, B., & Chatterjee, M. (2020). *Understanding the spatio-temporal pattern of COVID-19 outbreak in India using GIS and India's response in managing the Pandemic*. *Regional Science Policy & Practice*. doi:10.1111/rsp3.12359
- [2] Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). *GIS-based spatial modeling of COVID-19 incidence rate in the continental United States*. *Science of The Total Environment*, 138884. doi:10.1016/j.scitotenv.2020.138884 health indicators . PLoS ONE 14(2): e0210697. 2019
- [3] Yıldırım, M., & Güler, A. (2020). *COVID-19 severity, self-efficacy, knowledge, preventive behaviors, and mental health in Turkey*. *Death Studies*, 1–8. doi:10.1080/07481187.2020.1793434
- [4] Tantrakarnapa, K., Bhopdhornangkul, B., & Nakhaapakorn, K. (2020). *Influencing factors of COVID-19 spreading: a case study of Thailand*. *Journal of Public Health*. doi:10.1007/s10389-020-01329-5
- [5] Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., ... Agha, R. (2020). *World Health Organization declares Global Emergency: A review of the 2019 Novel Coronavirus (COVID-19)*. *International Journal of Surgery*. doi:10.1016/j.ijsu.2020.02.034
- [6] Cullen, W., Gulati, G., & Kelly, B. D. (2020). *Mental health in the Covid-19 pandemic*. *QJM: An International Journal of Medicine*. doi:10.1093/qjmed/hcaa110
- [7] Douglas, M., Katikireddi, S. V., Taulbut, M., McKee, M., & McCartney, G. (2020). *Mitigating the wider health effects of covid-19 pandemic response*. *BMJ*, m1557. doi:10.1136/bmj.m1557
- [8] Ratih Sari Wardani, Purwanto, Sayono and Aditya Paramananda(2019) *Clustering tuberculosis in children using K-Means based on geographic information system* doi:10.1063/1.5112483
- [9] Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). *GIS-based spatial modeling of COVID-19 incidence rate in the continental United States*. *Science of The Total Environment*, 138884. doi:10.1016/j.scitotenv.2020.138884
- [10] Ya Gao Ming Liu Yamin Chen Shuzhen Shi Jie Geng Jinhui Tian(2020). *Association between tuberculosis and COVID-19 severity and mortality: A rapid systematic review and meta-analysis* doi:10.1002/jmv.26311
- [11] Gao, Y., Liu, M., Chen, Y., Shi, S., Geng, J., & Tian, J. (2020). *Association between tuberculosis and COVID-19 severity and mortality: a rapid systematic review and meta-analysis*. *Journal of Medical Virology*. doi:10.1002/jmv.26311
- [12] Huang, I., Lim, M. A., & Pranata, R. (2020). *Diabetes mellitus is associated with increased mortality and severity of disease in COVID-19 pneumonia – A*

systematic review, meta-analysis, and meta-regression. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. doi:10.1016/j.dsx.2020.04.018

- [13] Srikanth Muthya, Renjith S Nair, Dr Thanga Prabhu Arokiaswamy, Anusha Prakash(2020)Using Machine Learning to assess Covid-19 risksi: doi:10.1101/2020.06.23.20137950
- [14] Marutho, D., Hendra Handaka, S., Wijaya, E., & Muljono. (2018). *The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News. 2018 International Seminar on Application for Technology of Information and Communication.* doi:10.1109/isemantic.2018.8549751
- [15] Ashutosh Kumar Dubey , Umesh Gupta , Sonal Jain(2018).Comparative Study of K-means and Fuzzy C-means Algorithms on The Breast Cancer Data international jurnal on advanced science engineering information technology.
- [16] Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. da F., & Rodrigues, F. A. (2019). *Clustering algorithms: A comparative approach.* PLOS ONE, 14(1), e0210236. doi:10.1371/journal.pone.0210236 .
- [17] Zhenyu Kang¹ , Shanshan Luo² , Yang Gui¹, Haifeng Zhou¹ Zili Zhang¹ , Chunxia Tian¹ , Qiaoli Zhou¹ ,Quansheng Wang¹ , Yu Hu² Heng Fan¹ ,Desheng Hu Obesity is a potential risk factor contributing to clinical manifestations of COVID-19 International Journal of Obesity doi:10.1038/s41366-020-00677-2
- [18] Yang, X., Zhao, J., Yan, Q., Zhang, S., Wang, Y., & Li, Y. (2020). *A case of COVID-19 patient with the diarrhea as initial symptom and literature review. Clinics and Research in Hepatology and Gastroenterology.* doi:10.1016/j.clinre.2020.03.013
- [19] Charlie H. Zhang, PhD1& Gary G. Schwartz, PhD, MPH, PhD2 Spatial Disparities in Coronavirus Incidence and Mortality in theUnited States: An Ecological Analysis as of May 2020 The Journal of Rural Health doi: 10.1111/jrh.12476
- [20] Aufa Qorina Putri,Fadhiila Senjaliana, Rini Aprilianty Riadi, Mochtar Hidayat, Nisa Ummaroh Fajarani Sasmita, Alfiani Yulita Sari, Edy Widodo.(2020). Pengelompokan provinsi terdampak pandemi covid-19 di Indonesia SEMINAR NASIONAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA (5th SENATIK).
- [21] Mohammad Khubeb Siddiqui , Ruben Morales-Menendez , Pradeep Kumar Gupta , Hafiz M.N. Iqbal , Fida Hussain , Khudeja Khatoon and Sultan ,Ahmad(2020)Correlation Between Temperature and COVID-19 (Suspected, Confirmed and Death) Cases based on Machine Learning Analysis Journal Pure Appl Microbiol doi:/10.22207/JPAM.14.SPL1.40
- [22] Dr. Viswa Chaitanya Chandu(2020)Identification of spatial variations in COVID-19 epidemiological data using KMeans clustering algorithm: a global perspective medRxiv doi:10.1101/2020.06.03.20121194;

- [23] Abdulkader Murad , and Bandar Fuad Khashoggi(2020)Using GIS for Disease Mapping and Clustering in Jeddah, Saudi Arabia International Jurnal of geo-information doi:10. 3390/i jgi9050328
- [24] Nuruddeen Musa Shanono,Frans Richard Kodong,Musheer Abdulwahid Abdo AL-Jaberi(2020) Monitoring Infectious Diseases Diffusion through GIS Journal of Science and Technology
- [25] Scarbrough, A. W., Hoffpauir, D. R., Hill, J., & Koppada, S. P. (2019). How mapping can reduce threats to public health. *International Journal of Health Promotion and Education*, 1–11. doi:10.1080/14635240.2019.1637767
- [26] C.Saranya, G.Manikandan (2013) A Study on Normalization Techniques for Privacy Preserving Data Mining International Journal of Engineering and Technology (IJET)
- [27] J. Liang and L. Jin, “Multi-perspective modeling of computer sales system Based on Unified Modeling Language,” *Proceedings of 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference, ITOEC 2020*, no. Itoec, pp. 109–113, 2020, doi: 10.1109/ITOEC49072.2020.9141934.
- [28] Ajay D. Shinde (2017) “ *EXTRACTING OBJECT MODEL INFORMATION FROM DATA FLOW DIAGRAM* ”*INTERNATIONAL JOURNAL OF RESEARCHES IN BIOSCIENCES, AGRICULTURE AND TECHNOLOGY*
- [29] STATMAT.(2020) Statistika Cara menghitung nilai rata-rata . April. <https://www.statmat.net/cara-menghitung-nilai-rata-rata/>
- [30] COVID-19 Government indonesia (2021),Penting Memahami Faktor Penyebab Lonjakan Agar Terhindar Gelombang Baru COVID-19.<https://covid19.go.id/p/berita/satgas-penting-memahami-faktor-penyebab-lonjakan-agar-terhindar-gelombang-baru-covid-19>
- [31] Webagus (2020) , Knowledge Discovery In Database (KDD) <https://www.webagus.id/2020/02/knowledge-discovery-in-database-kdd.html>
- [32] F. N. R. F. J. Aziz, B. D. Setiawan, and I. Arwani, (2018).“Implementasi Algoritma K-Means untuk Klasterisasi Kinerja Akademik Mahasiswa,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 6, pp. 2243–2251,
- [33] S. Sembiring, M. Zarlis, D. Hartama, R. S, and E. Wani, (2011). “Prediction Of Student Academic Performance By An Application Of Data Mining Techniques,” *International Conference on Management and Artificial Intelligence*, vol. 6, pp. 390–394,