

Analisis Pengelompokan Wilayah Penyebaran Covid-19 Di Indonesia Dengan Metode *Clustering* Menggunakan Algoritma *K-Means* Dan *K-Medoids*

Chandra Halim¹, Hindriyanto Dwi Purnomo², Teguh Wahyono³

Universitas Kristen Satya Wacana, Jl. Diponegoro 52-60, Salatiga 50711, Indonesia

Email: 672018305@student.uksw.edu¹, hindriyanto.purnomo@uksw.edu², teguh.wahyono@uksw.edu³

Abstrack - Corona virus is a disease that attacks human respiratory tract infections that are generally mild, such as flu and cough. If not treated quickly will result in death. This virus is quickly transmitted from human to human through the air and in contact. To reduce the spread of the virus, it requires clustering using the K-Means and K-Medoids algorithm, this method works to partition objects into groups. The clustering was obtained based on data on total cases, total deaths and total cures. Based on the results of this study, the K-Means algorithm is more optimal than the K-Medoids in clustering regions in Indonesia. It is proven that the best value of the Davies Bouldin Index from the K-Means algorithm is 0.158 with $k = 4$ and the K-Medoids algorithm is 0.806 with $k = 5$. The results of clustering are based on the most optimal value, namely the K-Means algorithm, showing cluster 1 Central Java and Java. East is at the top due to high case and death rates.

Keywords - *K-Means*, *K-Medoids*, *Covid-19*, *cluster*.

Intisari - Corona virus merupakan sebuah penyakit yang menyerang infeksi saluran pernafasan manusia yang umumnya ringan, seperti flu dan batuk. Jika tidak ada penanganan yang cepat akan mengakibatkan kematian. Virus ini dengan cepat menular ke manusia ke manusia melalui udara dan bersentuhan. Untuk mengurangi penyebaran virus dibutuhkan klastrisasi menggunakan algoritma *K-Means* dan *K-Medoids*, metode ini bekerja untuk mempartisi objek kedalam kelompok. Klastrisasi tersebut diperoleh berdasarkan data total kasus, total kematian dan total kesembuhan. Berdasarkan hasil dari penelitian ini, algoritma *K-Means* lebih optimal dari pada *K-Medoids* pada mengklasterisasi daerah-daerah di Indonesia. Dibuktikan pada nilai terbaik *Davies Bouldin Index* dari algoritma *K-Means* sebesar 0.158 dengan $k = 4$ dan algoritma *K-Medoids* sebesar 0.806 dengan $k = 5$. Hasil klasterisasi berdasarkan nilai yang paling optimal yaitu algoritma *K-Means*, memperlihatkan *cluster* 1 Jawa Tengah dan Jawa Timur menjadi yang teratas dikarenakan tingkat kasus serta tingkat kematian yang tinggi.

Kata Kunci - *K-Means*, *K-Medoids*, *Covid-19*, *cluster*.

I. PENDAHULUAN

Pada tanggal 7 Januari 2020, China mengidentifikasi kasus penyakit berjenis virus Corona atau *covid-19* [1]. Infeksi *covid-19* disebabkan oleh sekelompok virus yang menginfeksi sistem pernapasan manusia. Rata – rata kasus virus *covid – 19* menimbulkan gejala seperti menyebabkan infeksi, pernapasan yang berat seperti *pneumonia*, *Middle East Respiratory Syndrome* (MERS), dan *Severe Acute Respiratory Syndrome* (SARS) [2].

Kasus penyebaran virus *covid – 19* di dunia terus meningkat sampai menyebar ke Indonesia pada tanggal 2 Maret 2020. Menurut data dari Kementerian Kesehatan Republik Indonesia per tanggal 31 Maret 2021 dari 8.445.150 orang yang sudah diperiksa yang terkonfirmasi *covid-19* berjumlah 1.505.775, dengan jumlah kematian 40.754 dan sembuh 1.342.695 [3]. Hal ini tentunya berdampak ke masyarakat dari aspek pendidikan, ekonomi maupun sosial, oleh karena

itu pemerintah menerapkan sistem *lockdown*, isolasi mandiri selama 14 hari, dan penerapan protokol 3M (menjaga jarak, mencuci tangan, memakai masker) [4].

Sebagai upaya pencegahan perkembangan *Covid-19* di Indonesia, pada penelitian ini melakukan *clustering* untuk mengetahui daerah mana saja yang terdampak. Dengan pengelompokan ini juga dapat mengkategorikan kelompok daerah yang potensial penyebaran *Covid-19* dan kelompok daerah yang memiliki kasus *Covid-19* tertinggi. *Clustering* bertujuan pada pengelompokan seperti record, pengamatan serta membentuk kelas dari objek yang memiliki kemiripan dengan objek lainnya [1]. *Clustering* merupakan sebuah langkah awal yang dilakukan pada proses data mining yang menggunakan suatu metode analisis. Berbagai macam algoritma *clustering* yang dipakai peneliti sebelumnya seperti algoritma *K-Means*, *Improved K-Means*, *Fuzzy C-Means*, *DBSCAN*, *K-Medoids (PAM)*, *CLARANS* dan *Fuzzy*. Dari berbagai macam algoritma tersebut memiliki kekurangan dan kelebihan, namun memiliki prinsip yang sama yaitu mengelompokkan object sesuai dengan karakteristik dan mengukur jarak object kemiripan dalam kelompok maupun antar kelompok [5].

K-Means merupakan salah satu algoritma *clustering* yang masuk dalam gerombolan Unsupervised learning yang dipergunakan mengelompokkan data atau objek kedalam kelompok dengan sistem partisi [1]. Pengelompokan itu dilakukan jika ada object yang memiliki karakteristik yang sama maka akan dimasukkan ke dalam kelompok yang sama, dan objek yang karakteristik tidak sama akan dikelompokkan dalam kelompok yang tidak sama. Sedangkan algoritma *K-Medoids* adalah Analitis partisional, sebuah metode yang bertujuan untuk mendapatkan set *k-cluster* yang paling mendekati sebuah objek dalam melakukan pengelompokan sebuah data [5]. Algoritma *K-Means* dan *K-Medoids* merupakan pengelompokan iteratif yang set data partisinya tergantung dengan sejumlah *K cluster* yang sudah ditetapkan sejak awal [6].

Penelitian ini bertujuan untuk mencari klasterisasi terbaik dengan menggunakan algoritma *K-Means* dan *K-Medoids* dalam melakukan klasterisasi kasus *Covid-19* di Indonesia. Kedua metode itu dipilih karena sama - sama menggunakan metode *partitioning* [7]. Dataset yang di dapat pada penelitian ini adalah data public yang diperoleh dari Kemenkes. Penelitian ini menggunakan data total kasus, total kematian dan total sembuh dari kasus *Covid - 19* tiap provinsi yang ada di Indonesia. Berharap pada penelitian ini dapat bermanfaat dari berbagai kalangan dan memberikan wawasan mengenai pengelompokan penyebaran *Covid19*. Serta dapat membantu pemerintah dalam upaya mengurangi penyebaran *Covid-19* untuk memutus rantai penyebaran di Indonesia.

II. SIGNIFIKASI STUDI

A. Studi Literatur

Penelitian ini menggunakan artikel – artikel yang berhubungan dengan penelitian sebelumnya yang berkaitan dengan analisis *clustering*. Jurnal tersebut berjudul Analisis Algoritma *K-MEDOIDS CLUSTERING* dalam Pengelompokan Penyebaran *Covid-19* Di Indonesia. Dalam jurnal tersebut bertujuan untuk mengetahui seberapa banyak pengelompokan pada masalah kasus *Covid-19* di Indonesia menggunakan metode *K-Medoids Clustering*. Melonjaknya angka penyebaran wabah menjadi akar pokok permasalahan ini [5]. Hasil penelitian dari jurnal tersebut mendapatkan hasil yang baik, algoritma *K-Medoids* dapat melakukan pengelompokan data *covid-19* mana saja wilayah yang terinfeksi di wilayah masing – masing dengan pengklasteran terbaik dilakukan dengan tiga *cluster*. Penelitian pada jurnal tersebut memiliki kesamaan dalam proposal penelitian ini yaitu ingin mengelompokkan data wilayah *covid-19* dengan menggunakan metode *clustering*. Sedangkan perbedaannya adalah pada jurnal tersebut hanya menggunakan *K-Medoids* dalam melakukan *clustering* sedangkan

proposal penelitian ini menggunakan dua algoritma yaitu *K-Means* dan *K-Medoids* dalam melakukan *clustering*.

Penelitian ini didasari oleh artikel penelitian terdahulu yang berjudul Penerapan Algoritma *K-Means* Dalam Menentukan Tingkat Penyebaran Pandemi *Covid-19* Di Indonesia [8]. Penelitian tersebut sama - sama menggunakan algoritma *K-Means* dalam melakukan *Clustering*. Tetapi ada perbedaan, dalam jurnal tersebut terdapat perhitungan *Centroid* untuk mencari rata - rata setiap *cluster* kemudian diimplementasikan ke dalam *Rapidminer* untuk pengelompokan tiap wilayah. Sedangkan penelitian proposal ini menggunakan algoritma *K-Medoids* dan *Davies Bouldin Index* untuk menemukan validitas *cluster* terbaik.

Penelitian lain yang terkait adalah penelitian yang berjudul Analisis Data Mining Untuk *Clustering* Kasus *Covid-19* di Provinsi Lampung Dengan Algoritma *K-MEANS* [1]. Tujuan penelitian jurnal tersebut adalah untuk mengolah data *covid-19* provinsi Lampung menggunakan algoritma *K-Means* sebanyak 4 *cluster*. Hasil yang didapatkan dari jurnal tersebut, perhitungan manual memiliki hasil yang lebih baik jika dibandingkan dengan bantuan tools *RapidMiner*. Tetapi hasil dari kedua perhitungan tersebut sama – sama mendekati 0 yang dapat diartikan menghasilkan klaster yang baik. Kesamaan dalam proposal penelitian ini adalah menggunakan algoritma *K-Means* dalam melakukan klasterisasi. Sedangkan perbedaan pada penelitian ini yaitu menggunakan algoritma *K-Medoids* sebagai pembanding untuk menentukan algoritma yang lebih baik dalam kasus yang diperoleh.

B. Metode Penelitian

Tahapan Penelitian ini mengacu pada *AI Project Cycle* yang diambil dari buku *Artificial Intelligence Curriculum* [9]. Metode ini terdiri dari beberapa langkah yaitu *Problem scoping*, *Data Acquisition*, *Data Exploration*, *Modeling*, *Evaluation* yang dapat dilihat pada Gambar 1

AI Project Cycle merupakan langkah untuk memecahkan suatu permasalahan menggunakan metode ilmiah yang terbukti dan menarik kesimpulan mengenai hal itu. Tahapan dalam metode tersebut adalah sebagai berikut:

A. Problem Scoping

Definisi dari *Problem Scoping* adalah saat menyelesaikan suatu tujuan dari *system*, permasalahan tersebut harus di lingkup sesuai dengan *project* yang dikerjakan agar dapat mempermudah dalam menyelesaikan sebuah *project* [9]. Dalam *Problem Scoping* Terdapat 4W yang membantu dalam mengidentifikasi dan memahami masalah dengan cara yang lebih baik dan efisien. 4W dalam *Problem Scoping* antara lain:

1. Who

Who atau *Stakeholders* membantu dalam memahami dan mengkategorikan siapa semua yang terpengaruh secara langsung dan tidak langsung dengan masalah.

Kasus penelitian yang sedang dilakukan, masyarakat merasakan langsung dampak dari virus *Covid-19*. Sehingga membuat pemerintah mencari solusi untuk menanganinya. Dalam kasus ini secara tidak langsung tenaga pendidik membantu pemerintah dalam menangani *Covid-19*.

2. What

Bagian *What* membantu memahami dan mengidentifikasi sifat masalah dengan cara mengumpulkan bukti bahwa masalah yang dihadapi benar - benar ada. Seperti dari artikel surat kabar, Media, pengumuman dan masih banyak lagi.

Covid-19 sudah tersebar diseluruh dunia dan sampai ke Indonesia. Berita televisi, social media dan surat kabar memberitakan virus tersebut, dikarenakan masyarakat banyak terkena dampaknya serta terdapat banyak korban jiwa. Sehingga pada kasus penelitian ini mengambil data nyata dari total kasus *covid-19* di Indonesia. Data tersebut diambil dari Kemenkes.

3. *Where*

Where untuk mengetahui “dimana” masalah, situasi, dan lokasi. Kasus virus *Covid-19* pertama terjadi di China dan menyebar luas ke seluruh dunia dan sampai ke Indonesia. Akibatnya menyebabkan masyarakat panik karena virus menyebar dengan cepat tetapi obat belum tersedia.

4. *Why*

Why tentang manfaat yang akan diperoleh Stakeholders dari solusi dan bagaimana hal itu akan menguntungkan masyarakat.

Penelitian ini menggunakan solusi clustering pada daerah – daerah yang memiliki perkembangan Covid-19 yang tinggi. Fungsi dari cluster itu dapat membantu pemerintah dalam penanganan pertiap daerah. Jika diketahui cluster tersebut memiliki tingkat kasus Covid-19 yang tinggi, maka pemerintah bisa melakukan PPKM ataupun melakukan lockdown pada daerah tersebut. Jika daerah tersebut rawan atau memiliki potensi yang tinggi, pemerintah dapat memberi himbauan kepada masyarakat daerah itu untuk mematuhi protokol kesehatan.

B. *Data Acquisition*

Data Acquisition adalah proses pengumpulan data yang akurat agar dapat mengefisiensi system. Data dapat berupa teks, video, gambar, audio dan sebagainya. Data tersebut dikumpulkan dari berbagai sumber yang dapat dipercaya dan otentik seperti jurnal, surat kabar dan sebagainya.

Seperti pada penelitian yang sedang dilakukan saat ini, penelitian ini mengambil data dari *website* Kemenkes dan *Kaggle*. Data tersebut merupakan data mentah yang berbentuk tabel yang nantinya data tersebut akan diolah dan divisualisasikan.

C. *Data Exploration*

Data Exploration adalah proses mengatur data yang dikumpulkan secara seragam untuk pemahaman yang lebih baik. Data tersebut dapat disusun dengan cara visualisasi dan akan dapat menemukan tren dan pola darinya. Visualisasi dapat disusun dalam bentuk tabel, membentuk diagram atau membuat database.

Dalam penelitian yang sedang dilakukan, Data Exploration disusun dalam bentuk tabel yang di dapat dari data mentah kemenkes dan kaggle. Data tersebut disusun, diatur dan dipilih sedemikian rupa, sehingga data tersebut menjadi data matang yang dapat diolah dengan baik.

D. *Modeling*

Modeling digunakan untuk mengimplementasikan ide, melihat berbagai algoritma berkemampuan AI yang bekerja pada Computer Vision. Melalui beberapa model dan memilih model yang sesuai dengan kebutuhan. Setelah memilih model lalu menerapkannya.

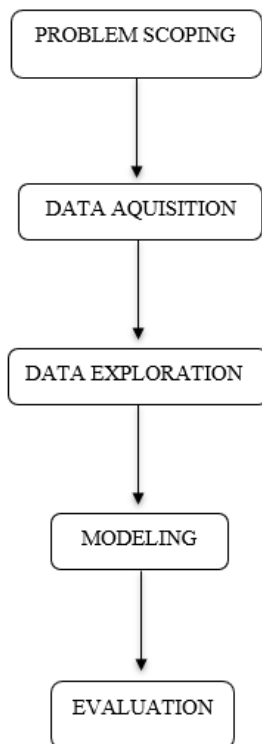
Penelitian ini menggunakan rapidminer sebagai modeling design. Terdapat banyak jenis modeling design yang ada pada rapidminer seperti Predictive, Segmentation, Associations, Correlations, Similarities, Feature Weights, Optimization, Time Series. Pada kasus penelitian ini menggunakan modeling segmentation algoritma K-means dan K-Medoids untuk membuat sebuah *cluster* [10].

E. *Evaluation*

Evaluation yaitu menguji dalam banyak cara, tahap pengujian model dikenal sebagai evaluasi. Dalam tahap ini, setiap model dapat dievaluasi, dicoba dan memilih model yang paling efisien sehingga mendapatkan hasil yang maksimal.

Pada penelitian tahap ini menggunakan tiga sampai tujuh kluster pada algoritma K-Means dan K-Medoids. Uji coba menggunakan acuan DBI (Davies Bouldin Index). Setelah

mendapatkan nilai DBI dari setiap cluster yang ada, akan dilakukan evaluasi yaitu dengan memilih model kluster yang efisien dan optimal. Pemilihan tersebut berdasarkan nilai DBI yang terkecil dikarenakan memiliki jarak intra-kluster yang minimum, sehingga di dalam kluster memiliki tingkat kesamaan karakteristik yang tinggi [11].



Gambar 1. Tahapan Penelitian.

III. HASIL DAN PEMBAHASAN

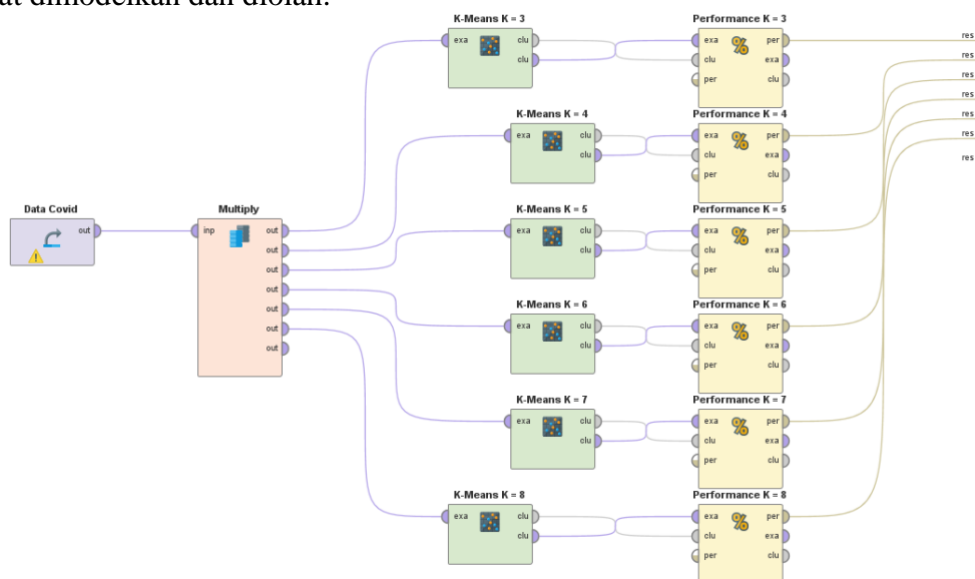
Pada tabel I merupakan data covid – 19 pada daerah – daerah di Indonesia. Data diperoleh dari Kemenkes, dalam data tersebut memiliki 4 atribut yaitu kota, total kasus, total kematian dan total sembuh. Data akan diolah ke dalam algoritma *K-Means* dan *K-Meodoids*. Kedua algoritma itu diolah sehingga mendapatkan nilai *centroid* dan DBI (*Davies Bouldin Index*). Nilai - nilai tersebut dibutuhkan untuk membuat sebuah *cluster* yang optimal.

TABEL I
DATA COVID-19

| Kota | Total Kasus | Total Kematian | Total Sembuh |
|--------------------|-------------|----------------|--------------|
| Aceh | 20140 | 862 | 15541 |
| Bali | 54079 | 1618 | 46181 |
| Banten | 61689 | 1457 | 50554 |
| Bengkulu | 11671 | 194 | 9530 |
| DKI Jakarta | 636383 | 9270 | 527060 |
| Yogyakarta | 72560 | 1892 | 54352 |
| Jambi | 13935 | 284 | 11892 |
| Jawa Barat | 440379 | 6008 | 348694 |
| Jawa Tengah | 285348 | 12562 | 223706 |
| Jawa Timur | 189705 | 13635 | 160296 |
| Kalimantan Barat | 16811 | 413 | 14371 |
| Kalimantan Selatan | 37176 | 1091 | 34774 |
| Kalimantan Tengah | 27808 | 556 | 21082 |

| Kota | Total Kasus | Total Kematian | Total Sembuh |
|---------------------------|-------------|----------------|--------------|
| Kalimantan Timur | 83762 | 2003 | 73864 |
| Kalimantan Utara | 14274 | 210 | 12318 |
| Kepulauan Bangka Belitung | 23188 | 365 | 20761 |
| Kepulauan Riau | 30637 | 639 | 23933 |
| Lampung | 24378 | 1149 | 19184 |
| Maluku | 10465 | 164 | 7708 |
| Maluku Utara | 6526 | 142 | 4650 |
| Nusa Tenggara Barat | 13466 | 498 | 11458 |
| Nusa Tenggara Timur | 22933 | 493 | 17388 |
| Papua | 21503 | 213 | 11696 |
| Papua Barat | 13087 | 199 | 9829 |
| Riau | 74899 | 2019 | 67986 |
| Sulawesi Barat | 6251 | 127 | 5595 |
| Sulawesi Selatan | 67280 | 1021 | 62671 |
| Sulawesi Tengah | 14669 | 423 | 12988 |
| Sulawesi Tenggara | 12447 | 252 | 10412 |
| Sulawesi Utara | 17424 | 572 | 15525 |
| Sumatera Barat | 55675 | 1259 | 48840 |
| Sumatera Selatan | 31147 | 1569 | 27088 |
| Sumatera Utara | 38040 | 1232 | 33773 |

Tabel yang terdapat pada gambar 1 dilakukan proses pengecekan missing value terlebih dahulu. Setelah tidak ada missing value data tersebut dapat dimasukkan ke dalam *rapid miner* agar dapat dimodelkan dan diolah.



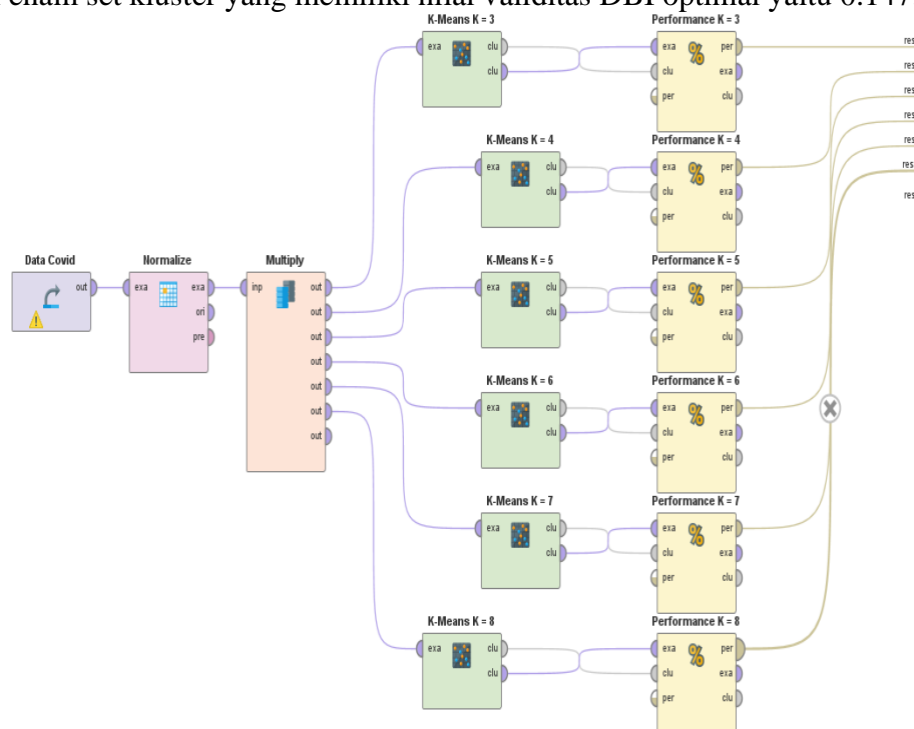
Gambar 2. Modeling desain K-Means

Pada gambar 2 merupakan modeling desain yang dilakukan tanpa melakukan normalisasi. Pada modeling desain gambar diatas, diawali dengan *Retrieve* yang digunakan untuk memasukan data berupa excel atau csv untuk dibaca oleh *Rapid Miner*. Selanjutnya dilakukan *Multiply* yang berguna untuk menyalin data yang ada pada *Retrieve book*. Setelah proses penyalinan data, data tersebut akan di kluster ke dalam *clustering K-Means* setiap kluster tersebut memiliki k yang berbeda – beda. Dilanjutkan dengan *cluster distance performance* yang digunakan untuk menghasilkan model kluster *centroid*, Set kluster, serta melakukan pengukuran *Davies Boudin Index*.

TABEL II
HASIL DBI K-MEANS TANPA NORMALISASI

| Jumlah cluster | Nilai DBI |
|----------------------|-----------|
| <i>K-Means</i> k = 3 | 0.429 |
| <i>K-Means</i> k = 4 | 0.238 |
| <i>K-Means</i> k = 5 | 0.273 |
| <i>K-Means</i> k = 6 | 0.147 |
| <i>K-Means</i> k = 7 | 0.258 |
| <i>K-Means</i> k = 8 | 0.278 |

Hasil model desain yang sudah diproses menghasilkan nilai DBI pada setiap klusternya. Hasil yang didapatkan seperti pada tabel I. Berdasarkan beberapa cluster yang sudah di uji coba, didapatkan enam set cluster yang memiliki nilai validitas DBI optimal yaitu 0.147.



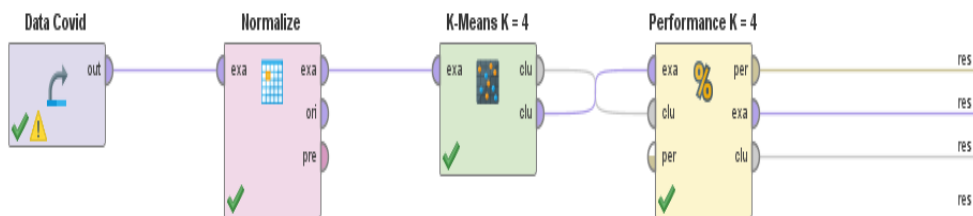
Gambar 3. Modeling desain normalisasi *K-Means*

Gambar 3 proses *clustering* menggunakan data yang ada pada *retrieve* kemudian dilakukan normalisasi. Normalisasi merupakan hal yang penting dalam melakukan *clustering*, perbedaan rentang nilai akan mempengaruhi hasil proses *clustering*. Oleh sebab itu untuk menyamakan rentang nilai menggunakan fungsi normalisasi [12].

TABEL III
HASIL DBI K-MEANS NORMALISASI

| Jumlah cluster | Nilai DBI |
|----------------------|-----------|
| <i>K-Means</i> k = 3 | 0.400 |
| <i>K-Means</i> k = 4 | 0.158 |
| <i>K-Means</i> k = 5 | 0.287 |
| <i>K-Means</i> k = 6 | 0.192 |
| <i>K-Means</i> k = 7 | 0.292 |
| <i>K-Means</i> k = 8 | 0.339 |

Dilihat tabel I dan II memiliki perbedaan nilai dan hasil nilai DBI optimal. Pada data yang sudah di normalisasi seperti tabel II, memiliki nilai DBI yang optimal dengan set *cluster* empat dan nilai DBI 0.158.



Gambar 4. Modeling visualisasi *K-Means*

Gambar 4 merupakan model desain kluster yang dipilih berdasarkan set kluster paling optimal. Kemudian performance example dan performance *cluster* model dihubungkan ke *result*. Hal itu dilakukan untuk mengetahui jumlah anggota setiap kluster serta visualisasinya dan nilai *centroid*.

TABEL IV
JUMLAH TIAP *CLUSTER K-MEANS*

| <i>Cluster</i> | Total |
|------------------|----------|
| <i>Cluster 0</i> | 29 items |
| <i>Cluster 1</i> | 2 items |
| <i>Cluster 2</i> | 1 items |
| <i>Cluster 3</i> | 1 items |

Set cluster terbagi menjadi set *cluster 0* yang terdiri dari 29 daerah, *cluster 1* terdiri dari 2 daerah, *cluster 2* dengan 1 daerah dan *cluster 3* memiliki 1 daerah.

TABEL V
NILAI CENTROID *K-MEANS*

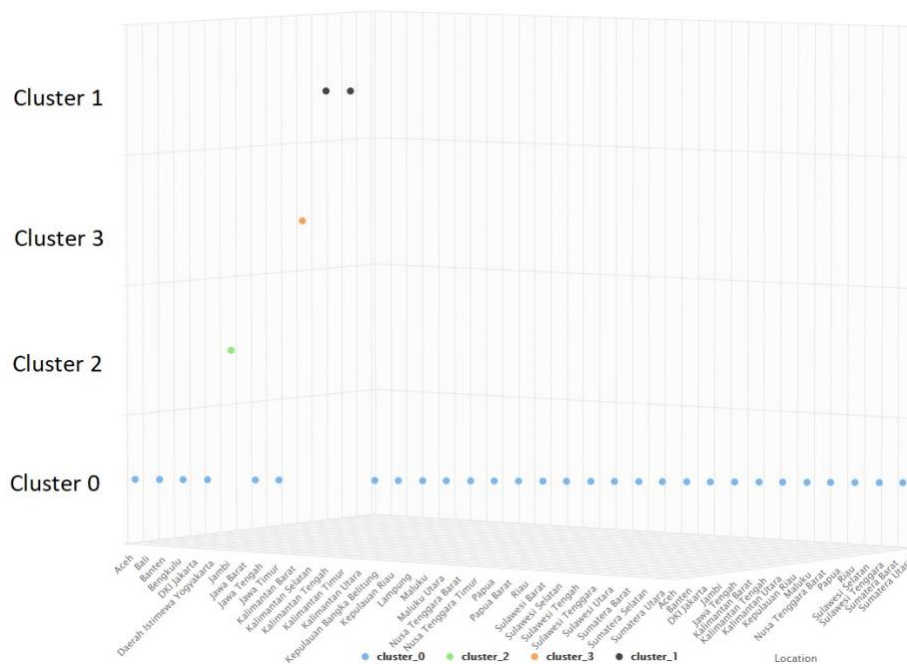
| Attribute | <i>Cluster 0</i> | <i>Cluster 1</i> | <i>Cluster 2</i> | <i>Cluster 3</i> |
|-----------------|------------------|------------------|------------------|------------------|
| Total Cases | -0.322 | 1.216 | 4.187 | 2.727 |
| Total Deaths | -0.342 | 3.284 | 2.156 | 1.195 |
| Total Recovered | -0.320 | 1.197 | 4.259 | 2.629 |

Tabel *centroid* menampilkan nilai – nilai *centroid* dalam masing masing *cluster*.

TABEL VI
JUMLAH TIAP *CLUSTER K-MEANS*

| <i>Cluster 0</i> | <i>Cluster 1</i> | <i>Cluster 2</i> | <i>Cluster 3</i> |
|--------------------|------------------|------------------|------------------|
| Aceh | Jawa Tengah | DKI jakarta | Jawa Barat |
| Bali | Jawa Timur | | |
| Banten | | | |
| Bengkulu | | | |
| Yogyakarta | | | |
| Jambi | | | |
| Kalimantan barat | | | |
| Kalimantan Selatan | | | |
| Kalimantan Tengah | | | |
| Kalimantan Timur | | | |

| Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 |
|---------------------|-----------|-----------|-----------|
| Kalimantan Utara | | | |
| Kepulauan Bangka | | | |
| Kepulauan Riau | | | |
| Lampung | | | |
| Maluku | | | |
| Maluku Utara | | | |
| Nusa Tenggara Barat | | | |
| Nusa Tenggara Timur | | | |
| Papua | | | |
| Papua Barat | | | |
| Riau | | | |
| Sulawesi Barat | | | |
| Sulawesi Selatan | | | |
| Sulawesi Tengah | | | |
| Sulawesi Tenggara | | | |
| Sulawesi Utara | | | |
| Sumatera Barat | | | |
| Sumatera Utara | | | |



Gambar 5. Visualisasi cluster K-Means k = 4

Tabel IV merupakan daerah – daerah di Indonesia yang sudah dikelompokkan berdasarkan nilai *centroid*. Kemudian di visualisasikan seperti pada gambar 5.

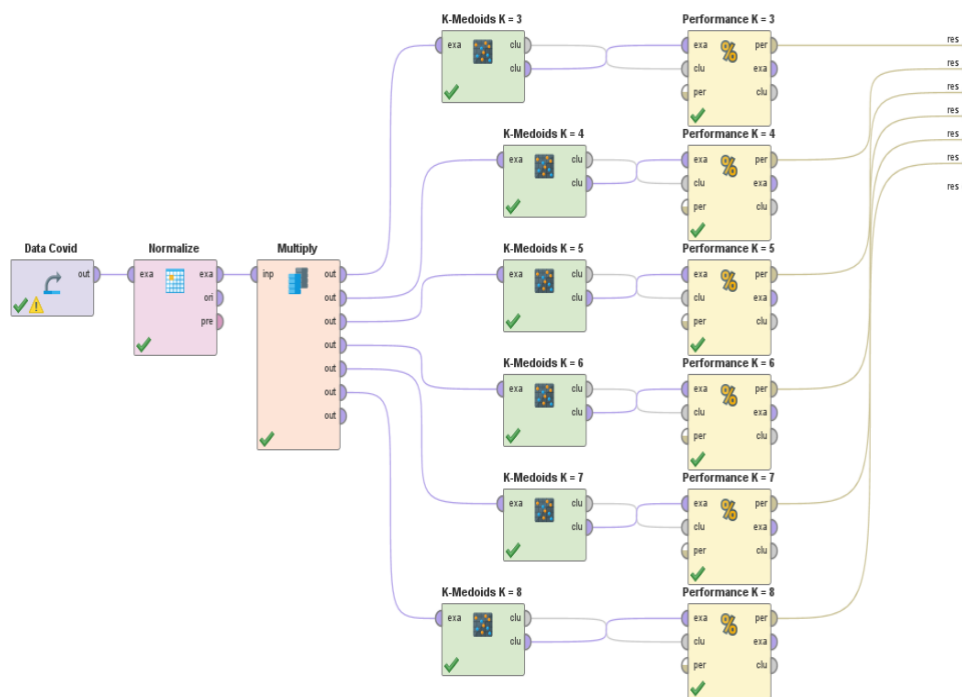
Berdasarkan hasil penelitian dan perhitungan menggunakan algoritma *K-Means*, setiap klusternya mendapatkan karakteristik. Kluster 1 memiliki karakteristik dengan total kematian yang paling besar, dilanjutkan dengan kluster 2 dan kluster 3 dengan karakteristik total kasus dan penyembuhan yang besar, kemudian pada kluster 0 merupakan daerah yang potensial dalam berkembangnya virus *Covid – 19*.

Dari karakteristik setiap *cluster*, dapat menjadi sebuah referensi pemerintah dalam melakukan pencegahan berkembangnya virus *Covid-19* di Indonesia. Pencegahan tersebut bisa

dilakukan dengan cara pada daerah kluster 1 dilakukan *lockdown* dikarenakan jumlah kematian yang besar, sehingga masyarakat berhenti beraktifitas sementara waktu untuk mencegah kematian yang semakin banyak. Diharapkan dengan aktifitas masyarakat yang berhenti sementara waktu dapat mengurangi dampak dari virus *Covid-19*.

Kemudian pemerintah bisa melakukan PPKM dan isolasi mandiri pada daerah yang ada pada kluster 3 dan kluster 2 dikarenakan total kasus dan kesembuhan yang seimbang. Diadakannya PPKM oleh pemerintah diharapkan jumlah kesembuhan lebih besar, dan kasus yang semakin kecil.

Selanjutnya pada daerah kluster 0 merupakan daerah yang rata - rata kasus *Covid-19* yang rendah. Sehingga pemerintah dapat memberikan sebuah himbauan dengan cara masyarakat harus memenuhi protokol 3M (menjaga jarak, mencuci tangan, memakai masker) ataupun dapat melakukan PPKM tingkat 1.



Gambar 6. Modeling desain normalisasi *K-Medoids*

Pada gambar 6 merupakan modeling design *K-Medoids*. Modeling design dilakukan dengan menginput data *covid – 19* yang kemudian di normalisasikan agar tidak ada perbedaan rentang nilai. Setelah dilakukan normalisasi, data akan dibuat enam bagian yang setiap bagiannya berisi nilai *K K-Medoids*. Setiap *cluster* akan diberikan *performance* untuk melihat nilai tiap *cluster* yang didapatkan.

TABEL VII
HASIL DBI *K-MEDOIDS* NORMALISASI

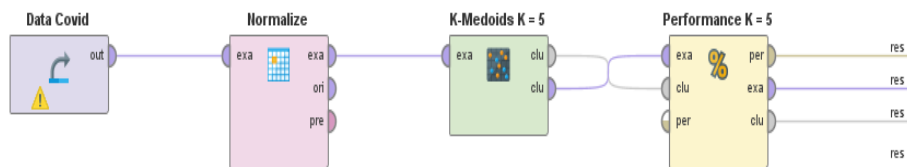
| Jumlah <i>cluster</i> | Nilai DBI |
|------------------------|-----------|
| <i>K-Medoids</i> k = 3 | 0.989 |
| <i>K-Medoids</i> k = 4 | 0.829 |
| <i>K-Medoids</i> k = 5 | 0.806 |
| <i>K-Medoids</i> k = 6 | 0.946 |
| <i>K-Medoids</i> k = 7 | 1.254 |
| <i>K-Medoids</i> k = 8 | 1.050 |

Dari modeling yang sudah dilakukan, mendapatkan hasil DBI seperti pada tabel VI. Nilai DBI yang paling optimal diambil dari nilai paling rendah diantara *cluster*_lainnya, sehingga didapatkan nilai 0.806 dengan $k = 5$.

TABEL VIII
HASIL DBI *K-MEDOIDS* NORMALISASI

| <i>Cluster</i> | Total |
|------------------|----------|
| <i>Cluster 0</i> | 11 items |
| <i>Cluster 1</i> | 9 items |
| <i>Cluster 2</i> | 2 items |
| <i>Cluster 3</i> | 2 items |
| <i>Cluster 4</i> | 9 items |

Lima *cluster* tersebut terdiri dari *cluster 0* dengan sebelas item, *cluster 1* dengan sembilan items, *cluster 2* dengan dua items, *cluster 3* dengan dua items dan *cluster 4* dengan sembilan items.



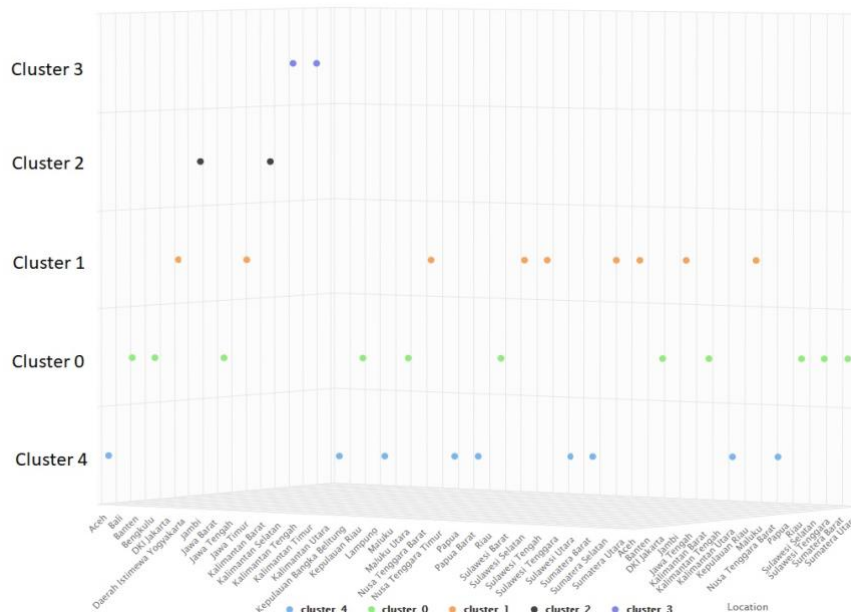
Gambar 7. Model untuk visualisasi *K-Medoids*

Cluster dengan nilai validitas optimal dilakukan pemodelan untuk menampilkan visualisasi serta nilai *centroid*.

TABEL IX
NILAI CENTROID *K-MEODIDS*

| Attribute | <i>Cluster 0</i> | <i>Cluster 1</i> | <i>Cluster 2</i> | <i>Cluster 3</i> | <i>Cluster 4</i> |
|-----------------|------------------|------------------|------------------|------------------|------------------|
| Total Cases | -0.270 | 0.460 | 2.727 | 0.860 | 0.423 |
| Total Deaths | -0.212 | 0.501 | 1.195 | 3.442 | 0.406 |
| Total Recovered | -0.250 | 0.463 | 2.629 | 0.907 | 0.416 |

Tabel diatas adalah nilai *centroid* yang berguna untuk membentuk sebuah *cluster* tetap.



Gambar 8. Visualisasi cluster K-Medoids

TABEL X
CLUSTER SETIAP DAERAH MENGGUNAKAN ALGORITMA K-MEDOIDS

| Cluster 0 | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|--------------------|-------------------|-------------|-------------|-------------------|
| Bali | Bengkulu | DKI Jakarta | Jawa Tengah | Aceh |
| Banten | Jambi | Jawa Barat | Jawa Timur | Kalimantan Barat |
| Yogyakarta | Kalimantan Utara | | | Kalimantan Tengah |
| Kalimantan Selatan | Maluku | | | Bangka |
| Kalimantan Timur | Maluku Utara | | | Riau |
| Lampung | Papua | | | NTB |
| Riau | Papua Barat | | | NTT |
| Sulawesi Selatan | Sulawesi Barat | | | Sulawesi Tengah |
| Sumatera Barat | Sulawesi Tenggara | | | Sulawesi Utara |
| Sumatera Selatan | | | | |
| Sumatera Utara | | | | |

Berdasarkan hasil penelitian dan perhitungan menggunakan algoritma *K-Medoids*. Dapat disimpulkan kluster – kluster tersebut memiliki sebuah karakteristik. Pada daerah kluster 3 memiliki karakteristik dengan tingkat kematian yang tinggi. Kluster 2 memiliki karakteristik dengan total kasus dan total kesembuhan yang tinggi. Kemudian kluster 1 mempunyai karakteristik kasus *Covid-19* tidak begitu tinggi, tetapi dari kasus tersebut lebih cenderung banyak kematian didapat daripada kesembuhan yang diperoleh. Pada kluster 4 mempunyai karakteristik yang kasus *Covid-19* tidak begitu tinggi, tetapi mempunyai tingkat kesembuhan yang baik Dan pada kluster 0 memiliki karakteristik pada kasus, kesembuhan dan kematian yang sangat rendah.

Karakteristik yang diperoleh dapat menjadi acuan pemerintah untuk menanggulangi *Covid-19*. Seperti pada daerah kluster 3, pemerintah dapat melakukan *lockdown* sementara dikarenakan tingkat kematian yang tinggi, Diharapkan dengan *lockdown* tersebut dapat mengurangi tingkat kematian serta mengurangi penyebaran virus *Covid-19*. Kemudian pada

daerah kluster 2, kluster 1 dan kluster 4 pemerintah dapat melakukan PPKM dengan skala bertingkat.

Pada daerah kluster 2 dapat menerapkan sistem PPKM tingkat 4 dikarenakan kasus yang tinggi. Sehingga lebih baik mengurangi aktifitas masyarakat tetapi tidak sampai memberhentikan total aktifitas. Pada daerah kluster 1 memiliki kasus *Covid-19* yang cukup rendah tetapi kasus kematian yang dominan sehingga dapat dilakukan PPKM tingkat 3, seperti masyarakat melakukan WFH, tempat ibadah membuka kapasitas berkisar 25% serta kegiatan belajar dilakukan secara daring.

Pada daerah kluster 4 memiliki tingkat kesembuhan yang baik dan kasus *Covid-19* yang cukup rendah. Sehingga pemerintah dapat melakukan PPKM tingkat 2 pada daerah tersebut. Dikarenakan daerah pada kluster 4 membaik karena tingkat kesembuhan dominan, daerah tersebut diperbolehkan 50% bekerja, belajar dan hal aktifitas lainnya dengan syarat sudah vaksin. Lalu pada kluster 0 merupakan daerah yang potensial untuk semakin meningkatnya kasus *covid - 19*. Pemerintah bisa memberikan peringatan atau PPKM tingkat 1, agar masyarakat tetap selalu waspada dan taat prokes yang sudah tersedia.

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan dari penelitian yang sudah dilakukan. Dapat disimpulkan, pengolahan data menggunakan algoritma *K-Means* memperoleh *cluster* terbaik dengan $k = 4$ dan nilai DBI 0.158. Kemudian pada algoritma *K-Medoids* memperoleh $k = 5$ dengan nilai DBI 0.806. Hasil pengujian validitas *cluster* terhadap klasterisasi *covid - 19* di Indonesia menggunakan algoritma *K-MEANS* dan *K-MEDOIDS*, pada penelitian ini *K-Means* lebih unggul daripada *K-Medoids* dengan memperoleh nilai DBI terbaik dan optimal yaitu 0.158 dengan $k = 4$. Sehingga pada kasus *covid - 19* lebih baik mengklasterisasi menggunakan algoritma *K-Means* dikarenakan jarak intra-kluster minimal dibandingkan dengan jarak inter-kluster *K-Medoids*. Sehingga objek dalam kluster memiliki tingkat kesamaan karakteristik yang tinggi.

Diharapkan dengan penelitian yang sudah dilakukan, pemerintah bisa mendapatkan manfaat dari penelitian ini mengenai menangani kasus *covid - 19*. Pemerintah bisa melakukan inovasi seperti klustering menggunakan algoritma *K-Means*, *K-Medoids* ataupun algoritma lainnya untuk mengelompokkan wilayah - wilayah yang darurat berdasarkan data yang didapat

Saran untuk penelitian lebih lanjut adalah untuk mengembangkan penelitian ini dengan melakukan perbandingan nilai validitas menggunakan *silhouette index*, atau pun menggunakan algoritma yang lainnya dalam menggambarkan klasterisasi.

REFERENSI

- [1] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *Jurnal Teknologi dan Sistem Informasi (JTISI)*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTISI>.
- [2] Z. I. Alfianti, "Pengelompokan Wilayah Penyebaran Covid-19 Di Kabupaten Karawang Menggunakan Algoritma K-Means," *Jurnal Ilmiah Informatika Komputer*, vol. 26, no. 2, pp. 111–122, 2021, doi: 10.35760/ik.2021.v26i2.4155.
- [3] Kemenkes, "Situasi Global," 2021. .
- [4] H. Harmawati and E. Yanti, "Kepatuhan Pengunjung Puskesmas Terhadap Protokol Kesehatan 3 M (Memakai Masker, Mencuci Tangan, Menjaga Jarak Dan Menghindari Kerumunan)," *Jurnal Abdimas Saintika*, vol. 3, no. 1, p. 94, 2021, doi: 10.30633/jas.v3i1.1106.

- [5] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, “Analisis algoritma K-Medoids clustering dalam pengelompokan penyebaran Covid-19 di Indonesia,” *Jti (Jurnal Teknologi Informasi)*, vol. 4, no. 1, pp. 166–173, 2020.
- [6] T. A. Cinderatama *et al.*, “Implementasi Metode K-Means , Dbscan , dan Meanshift Untuk Analisis Jenis Ancaman Jaringan Pada Intrusion Detection System,” *INOVTEK Polbeng*, vol. 7, pp. 169–184, 2022.
- [7] W. A. Triyanto, “Algoritma K-Medoids Untuk Penentuan Strategi Pemasaran Produk,” *Simetris : Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 6, no. 1, p. 183, 2015, doi: 10.24176/simet.v6i1.254.
- [8] N. Dwitri, J. A. Tampubolon, S. Prayoga, F. Ilmi Zer, and D. Hartama, “Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia,” *Jti (Jurnal Teknologi Informasi)*, vol. 4, no. 1, pp. 101–105, 2020.
- [9] M. Anita Karwal and M. Shweta Khurana, Eds., *ARTIFICIAL INTELLIGENCE CURRICULUM*, 9th ed. India: Intel, 2019.
- [10] S. S. Prasetyo, M. Mustafid, and A. R. Hakim, “Penerapan Fuzzy C-Means Kluster Untuk Segmentasi Pelanggan E-Commerce Dengan Metode Recency Frequency Monetary (Rfm),” *Jurnal Gaussian*, vol. 9, no. 4, pp. 421–433, 2020, doi: 10.14710/j.gauss.v9i4.29445.
- [11] S. S. Nober Six Salvanius Mendrofa, “Data Mining Untuk Penempatan Barang,” vol. 06, no. 04, 2022.
- [12] M. Tanzil Furqon and L. Muflikhah, “Clustering the Potential Risk of Tsunami Using Density-Based Spatial Clustering of Application With Noise (Dbscan),” *Journal of Enviromental Engineering and Sustainable Technology*, vol. 3, no. 1, pp. 1–8, 2016, doi: 10.21776/ub.jeest.2016.003.01.1.
- [13] Tedyyana, A., Fauzi, M., Enda, D., Ratnawati, F., & Syam, E. (2022). Perancangan Aplikasi Tanggap Api Berbasis Android Menggunakan Metode Design Sprint. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(2), 215-224. doi:<http://dx.doi.org/10.25126/jtiik.2022914022>