

PENERAPAN ALGORITMA *K-MEANS* DALAM MENENTUKAN TINGKAT PENYEBARAN PANDEMI *COVID-19* DI INDONESIA

Nayuni Dwitri , Jose A Tampubolon , Sandi Prayoga, P.P.P.A.N.W Fikrul Ilmi R.H Zer,
Dedy Hartama

Program Studi Teknik Informatika, STIKOM Tunas Bangsa Pematangsiantar

Jl. Jend. Sudirman Blok A, No. 1,2 dan 3, Kota Pematangsiantar, Sumatera Utara

nayunidwitri997@gmail.com, josetampubolon28@gmail.com, sandiprayoga00@gmail.com,
fikrulilmizer@gmail.com, dedyhartama@amiktunasbangsa.ac.id

Abstract - COVID-19 is an infectious disease caused by acute coronavirus 2 (severe acute respiratory coronavirus 2 or SARS-CoV-2) respiratory syndrome. The even distribution of COVID-19 cases in all provinces in Indonesia is a fairly rapid spread and harms all fields. The vast territory of Indonesia allows the need for grouping sections by region in Indonesia. This grouping will produce center points for the spread of COVID cases - 19. K-Means is one of the Clustering algorithms that is used to divide data into several groups with a system partition. This algorithm accepts input in the form of data without class labels. Due to the global pandemic that occurred many parties sought to participate in overcoming. This research was conducted for application in the spread of the co-19 pandemic in Indonesia. In this study, the K-Means algorithm is used to determine the level of co-19 distribution in regions in Indonesia.

Keywords - Algorithms, K-Means, Covid-19, The distribution, Pandemic

Abstrak - COVID-19 adalah penyakit yang dapat menular jika adanya kontak antara penular dengan orang lain, dan ditandai dengan gejala pada bagian pernapasan disebut SARS-CoV-2. Penyebaran kasus COVID-19 yang merata di seluruh provinsi di Indonesia, merupakan penyebaran yang cukup cepat dan berdampak negatif pada seluruh bidang. Luasnya wilayah Indonesia memungkinkan diperlukannya pengelompokan bagian bagian berdasarkan wilayah di Indonesia. Pengelompokan ini akan menghasilkan titik – titik pusat penyebaran kasus COVID -19. Salah satu algoritma *Clustering* adalah *K-Means* yang menggunakan beberapa kelompok untuk penempatan beberapa data dengan sistem partisi. Data-data tanpa label kelas diterima oleh Algoritma ini. Dikarenakan pandemi global yang terjadi banyak pihak berupaya ikut berperan serta dalam mengatasi. Penelitian ini dilakukan untuk penerapan dalam penyebaran pandemi covid-19 di Indonesia. Dalam penelitian ini menggunakan algoritma *K-Means* untuk menentukan bagaimana tingkat penyebaran covid-19 di daerah-daerah yang ada di Indonesia.

Kata kunci - Algoritma, K-Means , COVID-19, Penyebaran, Pandemi

I. PENDAHULUAN

COVID-19 adalah penyakit yang menular, dan ditandai oleh gejala pada bagian pernapasan akut (*coronavirus 2 (severe acute respiratory syndrome coronavirus 2* atau *SARS-CoV-2*). Virus ini merupakan keluarga besar Coronavirus yang dapat menyerang hewan. Penularan virus ini terjadi jika adanya kontak antar sesama manusia.. Ketika menyerang manusia, Penyakit yang disebabkan oleh virus ini adalah penyakit infeksi saluran pernafasan, seperti flu, MERS (*Middle East Respiratory Syndrome*), dan SARS (*Severe Acute Respiratory Syndrome*).[1]

Pada tanggal 2 Maret 2020 di Indonesia sebanyak dua kasus positif COVID-19 pertama kali dilaporkan. Terkonfirmasi berjumlah 1.528 kasus dan 136 kasus kematian pada 31 Maret 2020. Tingkat kematian COVID-19 di Indonesia sebesar 8,9%, adalah jumlah tertinggi di Asia Tenggara.[1]

Penyebaran kasus COVID-19 yang merata di seluruh provinsi di Indonesia, merupakan penyebaran yang cukup cepat dan berdampak negatif pada seluruh

bidang. Luasnya wilayah Indonesia memungkinkan diperlukannya pengelompokan bagian bagian berdasarkan wilayah di Indonesia. Pengelompokan ini akan menghasilkan titik- titik pusat penyebaran kasus COVID-19.

K-Means merupakan salah satu algoritma *Clustering* yang masuk dalam kelompok *Unsupervised learning* yang digunakan untuk mengelompokkan data kedalam beberapa kelompok dengan sistem partisi. Algoritma ini menerima masukan berupa data tanpa label kelas. Pada algoritma *K-Means*, komputer menerima data-data yang tidak diketahui kelasnya terkebih dahulu lalu mengelompokkannya. *Input* yang diterima ialah data dan jumlah kelompok (*cluster*) yang diinginkan.[2] Algoritma ini akan memasukan data kedalam beberapa kelompok tersebut.

II. METODOLOGI PENELITIAN

A. Tahap Pengumpulan Data

Data yang digunakan untuk penelitian ini adalah data jumlah penyebaran virus COVID-19 di

Indonesia yang diperoleh dari Kementerian Kesehatan Republik Indonesia. Data penyebaran virus COVID-19 yang digunakan pada penelitian ini adalah data penyebaran pada tanggal 9 Mei 2020.

B. COVID-19

COVID-19 adalah penyakit yang menular, dan ditandai oleh gejala pada bagian pernapasan akut (SARS-CoV-2). Virus ini merupakan keluarga besar Coronavirus yang dapat menyerang hewan. Penularan virus ini terjadi jika adanya kontak antar sesama manusia. Ketika menyerang manusia, Virus ini dapat menyebabkan penyakit infeksi saluran pernafasan, seperti MERS (Middle East Respiratory Syndrome), dan SARS (Severe Acute Respiratory Syndrome).[1] Virus ini telah tersebar lebih dari 190 negara lainnya. Oleh karena itu COVID-19 ditetapkan sebagai pandemi oleh WHO.

Sampai tanggal 8 Mei 2020, telah terjadi 3.672.238 kasus dan 245.045 kasus kematian di dunia. Sementara di Indonesia terdapat 12.776 kasus positif COVID-19 dan 930 kasus kematian. Dikarenakan pandemi global yang terjadi banyak pihak berupaya ikut berperan serta dalam mengatasi. Para dokter umum dan spesialis angkat bicara bersama guna memberi penjelasan singkat kepada masyarakat maupun imbauan agar menjaga kebersihan diri dan lingkungan sekaligus tak banyak keluar rumah.

C. Algoritma K-Means

Salah satu algoritma Clustering adalah K-Means yang digunakan untuk mengelompokkan data menjadi beberapa kelompok dengan beberapa Cluster.[2] Data – data dipilih menjadi beberapa kelompok dengan kriteria yang telah ditentukan lalu dikumpulkan menjadi satu dalam sebuah Cluster. Dimana setiap Cluster memiliki titik pusat yang disebut Centroid.

Berikut adalah tahapan-tahapan untuk melakukan optimasi menggunakan algoritma K-Means :[2]

- a) Pilihlah jumlah Cluster (k) yang diinginkan pada dataset
- b) Tentukan titik pusat (Centroid) secara acak/random pada tahap awal
- c) Hitunglah jarak terdekat setiap data dengan Centroid. Untuk menghitung jarak terdekat dengan Centroid adalah Euclidean distance (d). Dapat menggunakan rumus di bawah ini:

$$de = \sqrt{(xi - si)^2 + (yi - ti)^2} \dots\dots\dots(1)$$

Keterangan:
 (x ,y) = Koordinat objek
 (s ,t) = Koordinat Centroid
 i = Banyaknya objek

- d) Hitung kembali pusat Cluster dengan keanggotaan Cluster yang sekarang. Pusat Cluster adalah rata – rata dari semua data dalam sebuah Cluster. Dapat dihitung menggunakan rumus :

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \dots\dots\dots(2)$$

Keterangan:
 V_{ij} = Centroid rata – rata pada Cluster ke – i untuk variabel ke –j
 N_i = Jumlah anggota Cluster ke –i
 i,k = Indeks dari Cluster
 j = Indeks variabel
 X_{kj} = Nilai data ke –k variabel ke – j untuk Cluster tersebut

- e) Hitung kembali setiap objek menggunakan pusat Cluster baru (Centroid baru), ini merupakan tahap awal pembukaan iterasi baru. Jika anggota Cluster tidak mengalami perpindahan Cluster lagi, maka proses Clustering dinyatakan selesai. Tetapi, jika anggota Cluster mengalami perpindahan, maka kembali lagi pada langkah ke – c sampai anggota Cluster tidak mengalami perpindahan lagi.

III. HASIL DAN PEMBAHASAN

A. Perhitungan dengan Algoritma K-Means

Saat dilakukan perhitungan manual dengan data sampel sebanyak 34 provinsi yang terkena dampak penyebaran COVID-19 di Indonesia. Berikut adalah sampel data penyebaran COVID-19 di Indonesia per tanggal 9 Mei 2020:

Tabel 1. Sampel Data Penyebaran

No	Provinsi	Kasus Positif	Kasus Meninggal
1	Aceh	17	1
2	Bali	306	4
3	Banten	523	56
4	Bangka Belitung	28	1
5	Bengkulu	37	1
6	DI Yogyakarta	146	7
7	DKI Jakarta	5056	427
8	Jambi	62	0
9	Jawa Barat	1437	95
10	Jawa Tengah	959	66
11	Jawa Timur	1419	141
12	Kalimantan Barat	118	3
13	Kalimantan Timur	214	2
14	Kalimantan Tengah	189	7
15	Kalimantan Selatan	253	9
16	Kalimantan Utara	131	1
17	Kepulauan Riau	101	10
18	Nusa Tenggara Barat	330	6
19	Sumatera Selatan	278	6
20	Sumatera Barat	286	17
21	Sulawesi Utara	53	4
22	Sumatera Utara	179	21
23	Sulawesi Tenggara	71	2
24	Sulawesi Selatan	710	46
25	Sulawesi Tengah	75	3
26	Lampung	66	5
27	Riau	71	6
28	Maluku Utara	54	0

No	Provinsi	Kasus Positif	Kasus Meninggal
29	Maluku	32	2
30	Papua Barat	53	1
31	Papua	277	6
32	Sulawesi Barat	62	2
33	Nusa Tenggara Timur	12	0
34	Gorontalo	19	1

Pada tahap ini dapat diketahui bahwa :
 Jumlah Cluster : 3
 Jumlah Data : 34
 Jumlah Atribut :2 (Kasus Positif dan Kasus Meninggal)
 Iterasi ke -1

- Menentukan centroid awal secara acak.
 C_1 (Jawa Timur) = (1419,141) , C_2 (Nusa Tenggara Barat) = (330,6) , C_3 (Maluku) = (32,2)

Pada Cluster penyebaran COVID-19 yang menjadi penyebaran dalam jangka besar di ambil pada data ke 11 yaitu Jawa Timur,lalu untuk Cluster dengan penyebaran jangka menengah adalah data ke 18 yaitu Nusa Tenggara Barat,dan untuk Cluster dengan penyebaran jangka kecil adalah data ke 29 yaitu Maluku.

- Menghitung centroid terdekat.

Misalnya : Aceh (17,1)

ACEH (C1)	$\sqrt{(17-1419)^2+(1-141)^2}$	1408,972675
ACEH (C2)	$\sqrt{(17-330)^2+(1-6)^2}$	313,0399336
ACEH (C3)	$\sqrt{(17-32)^2+(1-2)^2}$	15,03329638

Langkah selanjutnya ialah menghitung jarak setiap data dengan Centroid dengan persamaan *Euclidean Distance*. Pada tahap ini jarak terdekat antara data dengan Cluster akan menentukan suatu data masuk ke dalam Cluster mana.Berikut adalah hasil perhitungan jarak pada Iterasi ke 1

Tabel 2.Hasil Perhitungan Iterasi 1

Provinsi	Jarak ke Centroid		
	Cluster 1	Cluster 2	Cluster 3
ACEH	1408,972675	313,0399336	15,03329638
BALI	1121,400018	24,08318916	274,0072992
BANTEN	900,0227775	199,3715125	493,9605247
BANGKA BEL.	1398,027539	302,0413879	4,123105626
BENGKULU	1389,073072	293,042659	5,099019514
DIYOGYAKARTA	1280,033203	184,0027174	114,1095964
DKI JAKARTA	3648,227652	4744,714638	5041,944169
JAMBI	1364,305684	268,0671558	30,06659276
JAWA BARAT	49,39635614	1110,571925	1408,074572
JAWA TENGAH	466,0740285	631,8552049	929,2066509
JAWA TIMUR	0	1097,335865	1393,947632
KALBAR	1308,298513	212,0212254	86,00581376
KALTIM	1212,990519	116,068945	182
KALTENG	1237,277657	141,0035461	157,0795977
KALSEL	1173,447911	77,0584194	221,1108319
KALUT	1295,586354	199,0628042	99,00505038
KEP.RIAU	1324,494243	229,0349318	69,46221995
NTB	1097,335865	0	298,0268444
SUMSEL	1148,958659	52	246,0325182
SUMBAR	1139,765327	45,35416188	254,4425279
SULUT	1372,852869	277,0072201	21,09502311
SUMUT	1245,79292	151,7432041	148,2228053
SULTRA	1355,147593	259,0308862	39

Provinsi	Jarak ke Centroid		
	Cluster 1	Cluster 2	Cluster 3
SULSEL	715,3362846	382,0994635	679,4262285
SULTENG	1351,066246	255,0176464	43,01162634
LAMPUNG	1359,818003	264,0018939	34,13209633
RIAU	1354,743149	259	39,20459157
MALUKU UTARA	1372,263094	276,0652097	22,09072203
MALUKU	1393,947632	298,0268444	0
PAPUA BARAT	1373,15549	277,0451227	21,02379604
PAPUA	1149,951738	53	245,0326509
SULBAR	1364,100436	268,0298491	30
NTT	1414,047383	318,0565987	20,09975124
GORONTALO	1406,982587	311,0401903	13,03840481

- Melakukan pengelompokkan berdasarkan cluster.Berikut adalah hasil pengelompokkan cluster pada Iterasi 1:

Tabel 3. Pengelompokkan Cluster Iterasi 1

Provinsi	Jarak Terdekat
ACEH	C3
BALI	C2
BANTEN	C2
BANGKA BEL.	C3
BENGKULU	C3
DIYOGYAKARTA	C3
DKI JAKARTA	C1
JAMBI	C3
JAWA BARAT	C1
JAWA TENGAH	C1
JAWA TIMUR	C1
KALBAR	C3
KALTIM	C2
KALTENG	C2
KALSEL	C2
KALUT	C3
KEP.RIAU	C3
NTB	C2
SUMSEL	C2
SUMBAR	C2
SULUT	C3
SUMUT	C3
SULTRA	C3
SULSEL	C2
SULTENG	C3
LAMPUNG	C3
RIAU	C3
MALUKU UTARA	C3
MALUKU	C3
PAPUA BARAT	C3
PAPUA	C2
SULBAR	C3
NTT	C3
GORONTALO	C3

- Menentukan Centroid baru untuk iterasi berikutnya dengan mencari rata-rata nilai dari setiap cluster.Berikut adalah nilai centroid baru untuk iterasi 2

Centroid	C1	2217,8	182,25
	C2	336,6	15,9
	C3	69,35	3,55

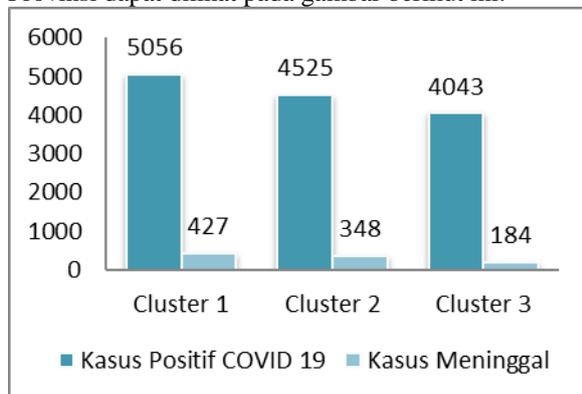
Jika tahapan iterasi telah mencapai hasil yang sama tanpa ada perpindahan lagi, maka perhitungan dihentikan.

Pada penelitian ini iterasi dilakukan sampai iterasi ke 6. Hasil akhir posisi *Cluster* dari *Centroid* iterasi ke 6 dapat dilihat pada table dibawah ini:

Tabel 4. Pembagian *Cluster* pada Iterasi ke -6

Provinsi	Letak <i>Cluster</i>
Aceh	C3
Papua Barat	C3
Banten	C3
Bangka Bel.	C3
Yogyakarta	C3
DKI Jakarta	C1
Maluku Utara	C3
Jawa Barat	C2
Jawa Tengah	C2
Jawa Timur	C2
Kalbar	C3
Kaltim	C3
Kalteng	C3
Kalsel	C3
Kalut	C3
Kep. Riau	C3
Bali	C3
Sumbar	C3
Sulut	C3
Sumut	C3
Sultra	C3
Sulsel	C2
Sulteng	C3
Lampung	C3
NTB	C3
Maluku	C3
Sumsel	C3
Papua	C3
Sulbar	C3
NTT	C3
Gorontalo	C3
Jambi	C3
Riau	C3

Hasil Pengolahan Data secara manual dengan 34 Provinsi dapat dilihat pada gambar berikut ini:



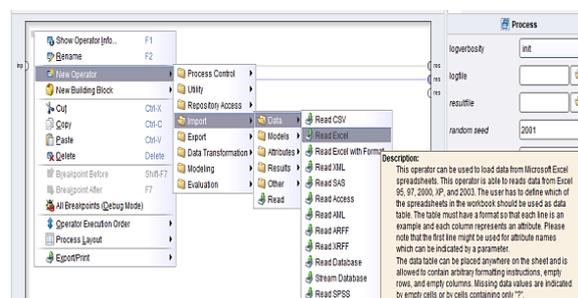
Gambar 1. *Cluster* tingkat penyebaran COVID-19

Hasil yang didapat dari grafik tersebut adalah *Cluster 1* sebagai penyebaran dalam jangka besar yaitu sebanyak 5056 kasus pasien positif COVID-19 dengan 427 kasus meninggal. Untuk *Cluster 2* sebagai penyebaran dalam jangka menengah yaitu dengan 4525 kasus positif COVID-19 dengan 348 kasus meninggal. Dan pada *Cluster 3* sebagai penyebaran jangka kecil yaitu dengan 4043 kasus positif COVID-19 dengan 184 kasus meninggal.

B. Implementasi Pada Aplikasi RapidMiner

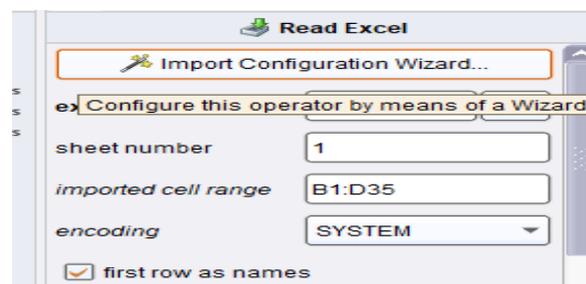
Dalam pengimplementasian pada aplikasi RapidMiner dapat dilihat sebagai berikut:

1. Import Data kedalam RapidMiner dalam bentuk sheet Excel



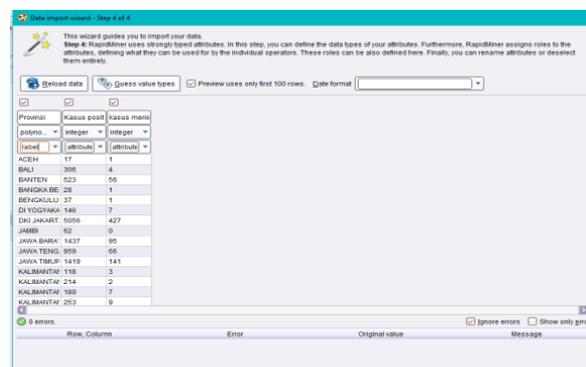
Gambar 2. Import Data

2. Import Configuration Wizard



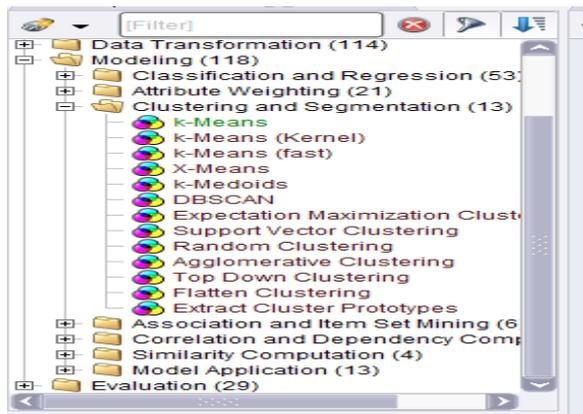
Gambar 3. Konfigurasi Wizard

3. Pilih File ,sesuaikan Anotasi dan Memilih Atribut dan Tipe Data



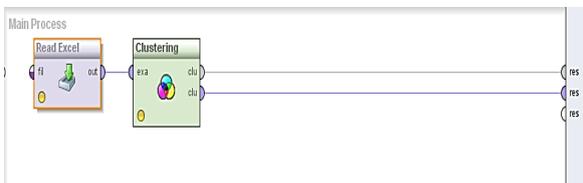
Gambar 4. Penyesuaian Atribut dan Tipe Data

4. Menambahkan Model K-Means



Gambar 5. Pemilihan Model K-Means

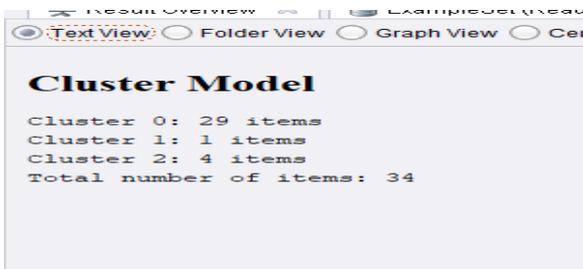
5. Menghubungkan Model K-Means



Gambar 6. Menghubungkan K-Means

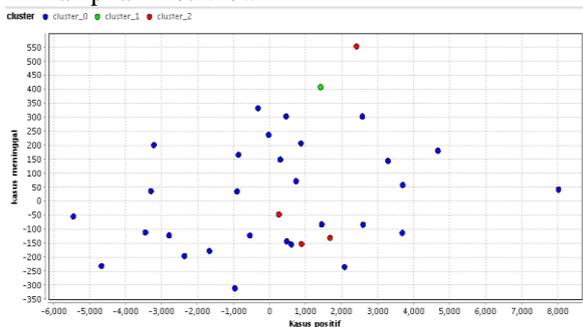
6. Jalankan Project

Tampilan Text View



Gambar 7. Hasil Tampilan Text View

Tampilan Plot View



Gambar 8. Hasil Tampilan Plot View

Setelah melakukan implementasi menggunakan Aplikasi RapidMiner ,hasil yang didapat setelah di implementasikan pada RapidMiner samadengan hasil pada perhitungan manual.

IV. KESIMPULAN

Berdasarkan penelitian dan hasil perhitungan menggunakan Algoritma K-Means dalam penyebaran COVID-19 di Indonesia,maka dapat ditarik kesimpulan sebagai berikut :

1. Pusat Cluster yang diperoleh yaitu untuk Cluster pertama berada pada provinsi DKI Jakarta.Pusat Cluster kedua berada pada provinsi Jawa Barat,dan pusat Cluster ketiga berada pada provinsi Banten
2. Pusat Cluster pertama merupakan daerah penyebaran COVID-19 terbesar di Indonesia dengan kasus positif dan kasus meninggal paling besar.
3. Daerah –daerah pada Cluster kedua dan ketiga merupakan penyebaran kasus COVID-19 yang cukup potensial,sehingga harus ,menjadi perhatian pemerintah untuk penanganannya

DAFTAR PUSTAKA

- [1] A.R. Setiawan, *Lembar Kegiatan Literasi Sainifik untuk Pembelajaran Jarak Jauh Topik Penyakit Coronavirus 2019 (COVID-19)*, Edukatif: Jurnal Ilmu Pendidikan Vol 2, No 1, 2020.
- [2] A. Wanto, et al, *Data Mining : Algoritma dan Implementasi*, Medan : Yayasan Kita Menulis, 2020 .
- [3] T. Naibaho, et al . *Kemampuan Siswa Menentukan Turunan Fungsi Aljabar Dengan Pembelajaran Model Kooperatif Tipe NHT (Kasus: Di Kelas XI MIA SMA Swasta Cerdas Bangsa Namurambe Tahun Pelajaran 2017/2018)*, Seminar Nasional Matematika dan Terapan, Vol. 1, 2019.
- [4] M.T.I. Rahmayani, *Analisis Clustering Tingkat Keparahan Penyakit Pasien Menggunakan Algoritma K-Means (Studi Kasus di Puskesmas Bandar Seikijang)*, JITI, Vol.1, No. 2, 2018.
- [5] A. Susilo, et al, *Coronavirus Disease 2019: Tinjauan Literatur Terkini . Jurnal Penyakit Dalam Indonesia*, Vol. 7, No. 1, 2020.
- [6] N. Meisida, et al. *K-Means untuk KLASIFIKASI PENYAKIT KARIES GIGI*, Kumpulan jurnal Ilmu Komputer (KLIK) Volume 01, No.01, 2014
- [7] A. Bastian, et al, *PENERAPAN ALGORITMA K-MEANS CLUSTERING ANALYSIS PADA PENYAKIT MENULAR MANUSIA (STUDI KASUS KABUPATEN MAJALENGKA)*. Jurnal Sistem Informasi (Journal of Information System), Volume 14, Issue 1, 2018.
- [8] F. Nasari, et al. *Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat*. Cogito Smart Journal, VOL. 2, NO. 2, 2016.