

Data Mining

Penerapan Algoritma K-means dalam Mengelompokkan Demam Berdarah Dengue Berdasarkan Kabupaten

Tresna Meida Mahening Tyas, Ade Irma Purnamasari

Program Studi Manajemen Informatika, STMIK IKMI Cirebon, Cirebon, Indonesia

INFORMASI ARTIKEL

Diterima Redaksi: 18 Februari 2023

Revisi Akhir: 06 Maret 2023

Diterbitkan Online: 07 Maret 2023

KATA KUNCI

Demam Berdarah Dengue, Data Mining, K-Means

KORESPONDENSI

Phone: +6283873354481

E-mail: meidatresna8@gmail.com

A B S T R A K

Demam berdarah *dengue* atau DBD adalah penyakit yang ditularkan oleh nyamuk *aedes aegypti* yang menyebabkan demam dan memunculkan pendarahan. Tercatat angka demam berdarah *dengue* mencapai 27.010 kasus dari Januari-Agustus 2022 di Jawa Barat berdasarkan Dinas Kesehatan Provinsi Jawa Barat. Hal ini menjadi penyebab lambatnya pencegahan maupun penanganan penyakit tersebut. Maka dari itu diperlukan pengelompokan penyebaran kasus demam berdarah *dengue* di Jawa barat. Pengelompokan menggunakan metode *K-means* serta data-data yang telah diperoleh dapat dikelompokkan dalam beberapa *cluster*, dimana proses pengelompokan *K-Means* menggunakan tools *RapidMiner*. Hasil yang diperoleh dari pengelompokan data menggunakan algoritma *K-Means* kasus penyebaran demam berdarah *dengue* berdasarkan kabupaten/kota di provinsi Jawa Barat dengan *cluster* optimal yaitu pada *cluster* 3 dengan nilai indeks *Davies Bouldin* sebesar 0.359. Adapun anggota *cluster* 0 berjumlah 100 anggota yang didominasi oleh Kabupaten Ciamis, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Kuningan dan Kabupaten Tasikmalaya. Sementara *cluster* 1 berjumlah 20 anggota yang didominasi Kota Bandung, serta *cluster* 2 berjumlah 96 anggota yang didominasi Kabupaten Bekasi, Kabupaten Bogor, Kabupaten Cirebon, Kabupaten Pangandaran, Kabupaten Sumedang, Kota Cimahi, dan Kota Tasikmalaya. Dari hasil ini diharapkan dapat membantu pemerintah untuk memetakan penyebaran penyakit demam berdarah *dengue* sebagai upaya penanganan penyebaran yang lebih luas serta dapat menjadi masukan kepada masyarakat untuk menjaga lingkungan melalui gerakan pemberantas nyamuk penularan demam berdarah *dengue*.

PENDAHULUAN

Demam berdarah *dengue* adalah infeksi virus yang ditularkan ke manusia melalui gigitan nyamuk *aedes aegypti* [1] Akibat dari demam berdarah suhu tubuh penderita menjadi sangat tinggi disertai dengan demam, muntah, sakit kepala, nyeri perut [2]. Dilaporkan pada tahun 2016 di daerah Pasifik Barat lebih dari 375.000 kasus, Filipina 176.411 dan Malaysia 100.028 kasus dengan dugaan demam berdarah. *World Health Organization* (WHO) mengatakan jika Indonesia menjadi negara dengan kasus demam berdarah tertinggi di wilayah Asia. Dalam data Kementerian Kesehatan Indonesia *Incidence Rate* (IR DBD) *dengue* pada tahun 2022 telah mencapai 34,33 persen per 100.000 penduduk [3]

Penyakit demam berdarah *dengue* dapat muncul sepanjang tahun dan dapat menyerang seluruh kelompok umur. Peran lingkungan dalam terjadinya penyakit memegang peranan penting dalam menentukan terjadinya proses interaksi antara manusia dan unsur penyebab dalam proses interaksi antara terjadinya penyakit, kondisi lingkungan yang buruk memberi keuntungan virus penyakit cepat berkembang biak. Penyebar penyakit DBD yaitu *aedes aegypti* menyukai lingkungan yang kurang sinar matahari, lembap, gelap dan dekat dengan sungai [4]

Demam Berdarah *dengue* adalah penyakit yang menjadi masalah Kesehatan masyarakat di setiap Kabupaten pada provinsi Jawa Barat. Menurut data Dinas Kesehatan Provinsi Jawa Barat Januari-Agustus 2022, tercatat 27.010 kasus demam berdarah di Jawa Barat, dengan 241 kasus di antaranya berakhir dengan kematian. Besarnya angka kasus penyebaran demam berdarah *dengue* mencerminkan permasalahan masih belum tertangani dengan baik. Maka dari itu, diperlukan cara pengelompokan atau klusterisasi penyakit DBD berdasarkan jumlah kasus yang banyak terjadi agar dapat membantu pemerintah dalam mengatasi angka penyebaran demam berdarah *dengue* secara cepat.

Pada penelitian sebelumnya yang telah dilakukan oleh [4] dengan judul “Data Mining: Penerapan Rapidminer dengan *K-Means Cluster* pada daerah Terjangkit Demam Berdarah *Dengue* (DBD) Berdasarkan Provinsi” yang melakukan pengelompokan penyakit DBD berdasarkan provinsi tahun 2014 – 2016 dengan menggunakan algoritma *k-means clustering* dimana hasil pemodelan terbagi menjadi 3 *cluster*, dimana *cluster* 1 memiliki kategori tinggi, *cluster* 2 dengan kategori tingkat sedang dan *cluster* 3 dengan kategori tingkat rendah. Penelitian ini menjadi salah satu referensi penulis dalam melaksanakan penelitian sehingga penulis dapat mengkaji teori yang digunakan dalam menganalisis penelitian yang dilakukan.

Oleh karena itu, penelitian ini melakukan pengelompokan penyakit DBD di Provinsi Jawa Barat berdasarkan kabupaten/kota dengan menggunakan algoritma *k-means* untuk memperoleh hasil dengan *cluster* optimal serta untuk mengetahui kabupaten/kota mana yang memiliki kasus penyebaran DBD mendominasi dari setiap *cluster* yang didapatkan. Sehingga pemerintah dapat menangani kabupaten/kota dengan kasus penyebaran dan dapat memberikan kebijakan pola hidup sehat kepada masyarakat mengenai kabupaten yang terjangkit DBD.

TINJAUAN PUSTAKA

Demam Berdarah Dengue

Demam Berdarah *Dengue* (DBD) ialah suatu penyakit yang disebabkan oleh virus *Dengue* yang menyebabkan demam dan memunculkan pendarahan. Gejala spesifik dari DBD adalah pendarahan pada kulit atau tanda pendarahan lainnya kadang terjadi di akhir waktu penyakit. Tragisnya jika penyakit ini didiagnosis dengan lambat dapat berujung kematian. Tragisnya jika penyakit ini didiagnosis dengan lambat dapat berujung kematian. Peristiwa demam berdarah cenderung mengalami peningkatan dan menyebar dengan cepat [1] Faktor penyebaran Demam Berdarah *Dengue* yaitu perubahan iklim dan kepadatan penduduk. Perubahan iklim memicu kondisi faktor lingkungan seperti curah hujan serta kelembapan suhu udara akibatnya nyamuk *aedes aegypti* berkembangbiak semakin cepat. Angka kepadatan penduduk yang semakin besar berdampak pada semakin banyak penduduk yang menempati daerah tersebut, sehingga peluang nyamuk *aedes aegypti* menggigit semakin besar [5]

Data Mining

Data mining ialah penemuan informasi baru dengan memeriksa pola ataupun aturan tertentu dalam data jumlah besar. Data mining diartikan sebagai proses menemukan pola-pola dalam data. Pola yang ditemukan memberikan keuntungan, umumnya keuntungan secara ekonomi. [6]

Knowledge Discovery in Database

Knowledge discovery In Database merupakan serangkaian proses yang tidak mudah untuk memeriksa serta mengetahui pola dalam suatu data, dimana pola itu bersifat asli, kemudian dapat dipahami dan bermanfaat untuk pengguna [7].

Clustering

Metode *Clustering* adalah salah satu metode dari fungsi data mining, algoritma *clustering* yaitu algoritma pengelompokan data menjadi *cluster* data tertentu [8] *Clustering* mengelompokkan sejumlah data ke dalam *cluster* baik pada tingkat kesamaan yang tinggi antar data di suatu *cluster*, begitupun sebaliknya data di dalam *cluster* tingkat ketidaksamaan tinggi antar *cluster* lainnya [9]

K-Means Clustering

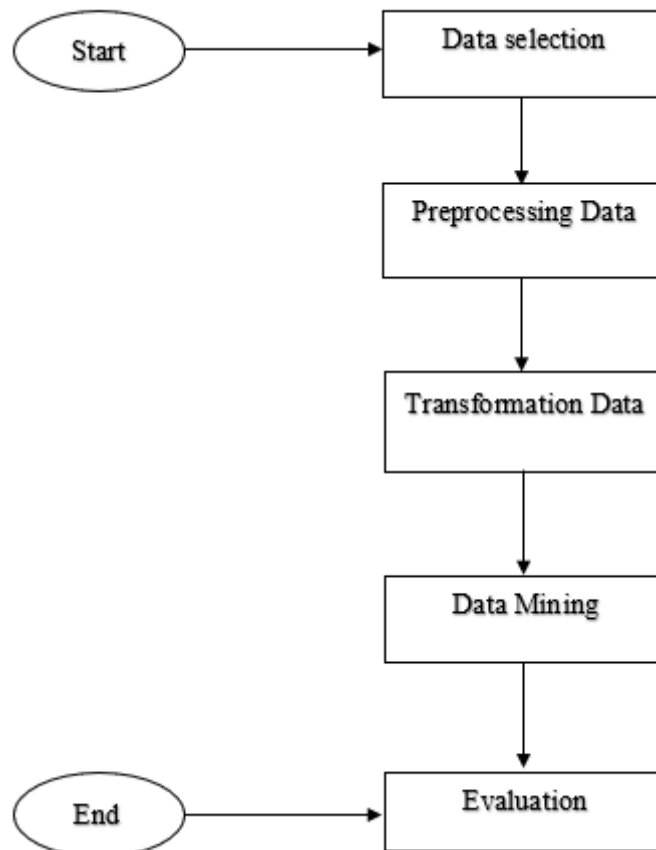
Algoritma *K-Means* yaitu algoritma yang termasuk *unsupervised learning* digunakan untuk pengelompokan data ke dalam beberapa cluster dengan sistem pemisah. *Unsupervised learning* adalah algoritma data mining untuk menemukan pola dari variabel (atribut), variabel (atribut) yang menjadi label tidak ditentukan [10].

Rapidminer

RapidMiner yaitu perangkat lunak untuk pengolahan data. *RapidMiner* memudahkan pengguna melakukan perhitungan data yang banyak dengan menggunakan beberapa operator. Hasil yang diperoleh dari *RapidMiner* bisa diperlihatkan secara tampilan dengan grafik, sehingga *tools RapidMiner* ini menjadi salah satu pilihan terbaik perangkat lunak untuk melakukan pengolahan data dengan teknik-teknik data mining [11].

METODOLOGI

Pada metodologi penelitian terdapat susunan kerangka kerja yang akan diikuti. Adapun kerangka kerja yang dilakukan pada penelitian ini adalah langkah-langkah *Knowledge Discovery in Database* atau biasa disebut dengan KDD, dapat dilihat pada *flowchart* berikut ini.



Gambar 1 *Flowchart* Tahapan KDD

HASIL DAN PEMBAHASAN

Data Selection

Data selection merupakan tahapan pertama dengan melakukan pemilihan atribut pada dataset. Dataset awal terdiri dari 8 atribut yaitu id, kode provinsi, nama provinsi, kode kabupaten/kota, nama kabupaten/kota, jumlah kasus, satuan, tahun.

Tabel 1 Dataset Demam Berdarah *Dengue*

id	Kode Provinsi	Nama Provinsi	Kode Kabupaten/Kota	Nama Kabupaten/Kota	Jumlah Kasus	Satuan	Tahun
1	32	Jawa Barat	3021	Kabupaten Bogor	1.834	Orang	2014
2	32	Jawa Barat	3022	Kabupaten Sukabumi	704	Orang	2014
3	32	Jawa	3023	Kabupaten Cianjur	339	Orang	2014

4	32	Barat Jawa	3024	Kabupaten Bandung	995	Orang	2014
5	32	Barat Jawa	3025	Kabupaten Garut	449	Orang	2014
...	...	...	...	...	...	...	...
215	32	Barat Jawa	3278	Kabupaten Tasikmalaya	909	Orang	2021
216	32	Barat Jawa	3279	Kabupaten Banjar	29	Orang	2021

Atribut yang dipilih dalam penyebaran kasus demam berdarah *dengue* adalah atribut nama kabupaten/kota, jumlah kasus serta tahun yang akan diproses pada *RapidMiner*.

Tabel 2 Hasil *Data Selection*

Nama Kabupaten/Kota	Jumlah Kasus	Tahun
Kabupaten Bogor	1.084	2014
Kabupaten Sukabumi	704	2014
Kabupaten Cianjur	339	2014
Kabupaten Bandung	995	2014
Kabupaten Garut	449	2014
...	...	...
Kabupaten Bekasi	1.844	2021
Kabupaten Depok	3.155	2021
Kabupaten Cimahi	590	2021
Kabupaten Tasikmalaya	909	2021
Kabupaten Banjar	29	2021

Preprocessing Data

Tahap *preprocessing* dilakukan pembersihan data dimana tidak ada *missing value* yang dapat menghambat proses data mining. Dapat dilihat bahwa atribut-atribut ini tidak memiliki *missing value* dan dapat melanjutkan proses pada tahap berikutnya.

Name	Type	Missing	Filter (4 / 4 attributes):
Id Nama kabupaten kota	Polynomial	0	Least KOTA TASIKMALAYA (7) Most KABUPATEN BANDUNG...
Cluster cluster	Nominal	0	Least cluster_1 (20) Most cluster_0 (100)
Jumlah kasus	Real	0	Min -0.928 Max 4.353
Tahun	Real	0	Min -1.524 Max 1.524

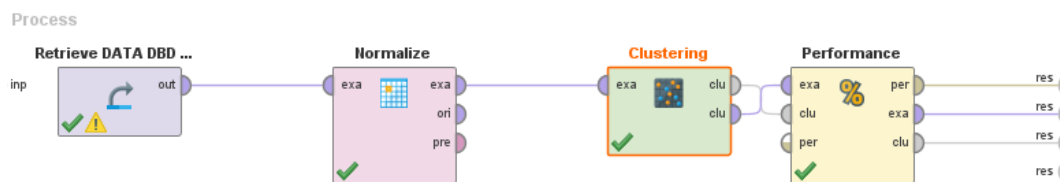
Gambar 2 Pengecekan *Missing Value*

Data Transformation

Tahap *data transformation* merupakan tahap dimana dataset dilakukan transformasi data dengan membuat format dataset baru dari dataset yang telah dipilih sebelumnya. Pada penelitian ini, dataset tidak ditransformasi sehingga dapat melanjutkan proses pada tahap berikutnya yaitu tahap data mining.

Data Mining

Berikutnya tahap data mining yang dilakukan menggunakan algoritma *K-Means* dengan memakai tools *RapidMiner*. Pada proses *RapidMiner* menggunakan beberapa Operator *RapidMiner* yaitu *retrieve data*, *normalize*, *k-means clustering*, dan *performance* yang dapat dilihat pada Gambar 3.

Gambar 3 Konfigurasi *RapidMiner*

Untuk operator *Retrieve* dapat mengakses informasi yang disimpan di Repositori dan memuatnya ke dalam proses. Operator ini dapat menyimpan data di dalam repositori dan memberikan informasi tambahan tentang objek *Rapidminer* yang diambil. Proses operator *Retrieve* dimulai dari mengimport data yang telah disimpan pada *file document*. Dalam operator *Clustering* yang menggunakan algoritma *K-Means*, *max runs* yang dipilih adalah 10, sedangkan untuk parameter *measure type* menjadi *NumericalMeasure* serta *numerical measure* menjadi *EuclideanDistance* untuk standar jarak antar titik *cluster*.

Evaluation

Setelah tahap data mining, selanjutnya tahap evaluasi dengan melihat kualitas dari cluster yang telah dibentuk. Dari proses operator *performance* diperoleh indeks *Davies Bouldin* sebagai berikut.

Tabel 3 Hasil DBI

Cluster	DBI
2	0.505
3	0.359
4	0.379
5	0.379
6	0.393
7	0.368
8	0.410
9	0.381
10	0.391

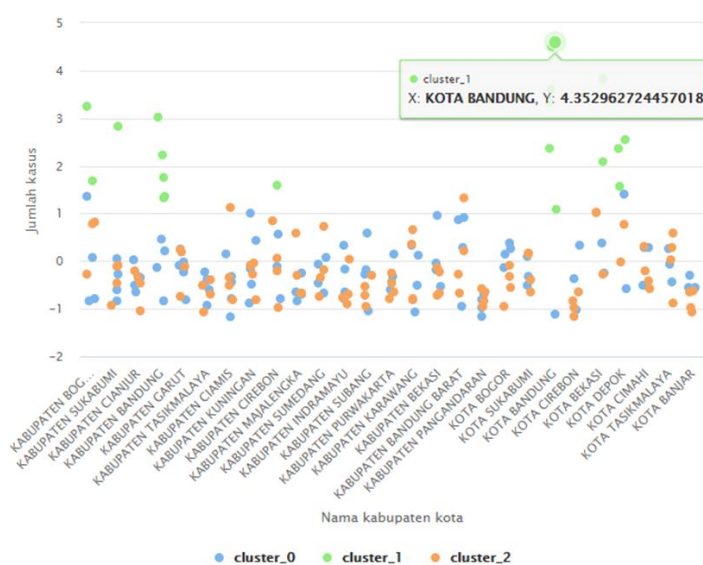
Pengelompokan yang menghasilkan kumpulan *cluster* dengan indeks *Davies Bouldin* terkecil dianggap sebagai *cluster* terbaik berdasarkan kriteria ini. Maka dari itu hasil pengelompokan data menggunakan algoritma *K-Means* kasus penyebaran demam berdarah *dengue* berdasarkan kabupaten/kota di provinsi Jawa Barat dengan melakukan uji coba sebanyak 10 kali didapatkan hasil *cluster* optimal yaitu pada $K=3$ dengan nilai indeks *Davies Bouldin* sebesar 0.359.

Cluster Model

```
Cluster 0: 100 items
Cluster 1: 20 items
Cluster 2: 96 items
Total number of items: 216
```

Gambar 4 *Cluster Model*

Dari *Cluster Model* seperti gambar di atas, dapat diketahui *cluster* yang didapat yaitu 3 *cluster* diantaranya *cluster 0* (C0), *cluster 2* (C2), *cluster 3* (C3). Pada *ExampleSet* dapat dilihat beberapa tampilan hasil *cluster*, yaitu *Plot View*. *Plot View* merupakan tampilan hasil *cluster* data secara keseluruhan sesuai dengan data yang telah diinputkan. Tampilan hasil *cluster* dapat dilihat pada gambar berikut.

Gambar 5 Tampilan *scatter*

Untuk hasil *clustering* menggunakan *RapidMiner*, dapat melihat tampilan *scatter* dari masing-masing cluster yang didapatkan dengan melihat warna-warna yang berbeda. Dari penjelasan tersebut diperoleh beberapa informasi bahwa dari 216 data yang dikelompokkan ke dalam 3 *cluster*. Adapun anggota *cluster* 0 yang berwarna biru sebanyak 100 anggota yang didominasi oleh Kabupaten Ciamis, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Kuningan dan Kabupaten Tasikmalaya. Sementara *cluster* 1 yang berwarna hijau sebanyak 20 anggota yang didominasi Kota Bandung, serta *cluster* 2 yang berwarna jingga sebanyak 96 anggota yang didominasi Kabupaten Bekasi, Kabupaten Bogor, Kabupaten Cirebon, Kabupaten Pangandaran, Kabupaten Sumedang, Kota Cimahi, dan Kota Tasikmalaya.

KESIMPULAN DAN SARAN

Berdasarkan hasil dan pembahasan pada bab-bab sebelumnya, kesimpulan yang di dapat dari penelitian ini yaitu, Hasil pengelompokan penyebaran Demam Berdarah Dengue berdasarkan kabupaten/kota menggunakan algoritma K-Means yaitu anggota cluster 0 yang berwarna biru sebanyak 100 anggota yang didominasi oleh Kabupaten Ciamis, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Kuningan dan Kabupaten Tasikmalaya. Sementara cluster 1 yang berwarna hijau sebanyak 20 anggota yang didominasi Kota Bandung, serta cluster 2 yang berwarna jingga sebanyak 96 anggota yang didominasi Kabupaten Bekasi, Kabupaten Bogor, Kabupaten Cirebon, Kabupaten Pangandaran, Kabupaten Sumedang, Kota Cimahi, dan Kota Tasikmalaya. Kabupaten tersebut diperlukan penanganan penyakit Demam Berdarah Dengue yang lebih serius oleh pemerintah. Hasil pengujian dari K optimal yang didapatkan menggunakan *RapidMiner* berdasarkan data penyebaran Demam Berdarah Dengue di Jawa Barat adalah K=3 dengan nilai indeks Davies Bouldin sebesar 0.359 mendekati 0.

Berikut ada beberapa saran yang dapat disampaikan, Pada penelitian ini masih memiliki kekurangan diharapkan untuk tugas akhir selanjutnya data yang digunakan dengan jumlah yang lebih banyak agar dapat menghasilkan data yang lebih akurat. Pada penelitian selanjutnya dapat dilanjutkan menggunakan algoritma data mining yang lain ataupun menggunakan dua algoritma yang berbeda sehingga hasil pengelompokan data lebih variatif.

DAFTAR PUSTAKA

- [1] E. M. Sembiring, "Penerapan K-Means Clustering Untuk Pengelompokan Penyebaran Demam Berdarah Dengue (DBD) Di Kabupaten Deli Serdang," *TIN: Terapan Informatika Nusantara*, vol. 2, no. 11, hlm. 673–677, Apr 2022, doi: 10.47065/tin.v2i11.1503.
- [2] S. H. Widiastuti dan R. Jumardi, "Pengelompokan Daerah Rawan Demam Berdarah dengan Metode K-Means Clustering," *Jurnal Informasi dan Teknologi*, Agu 2022, doi: 10.37034/jidt.v4i4.213.
- [3] D. Amelia, T. N. Padilah, dan A. Jamaludin, "Optimasi Algoritma K-Means Menggunakan Metode Elbow dalam Pengelompokan Penyakit Demam Berdarah Dengue (DBD) di Jawa Barat," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 11, hlm. 207–215, 2022, doi: 10.5281/zenodo.6831380.

- [4] K. Fatmawati dan A. P. Windarto, "DATA MINING: PENERAPAN RAPIDMINER DENGAN K-MEANS CLUSTER PADA DAERAH TERJANGKIT DEMAM BERDARAH DENGUE (DBD) BERDASARKAN PROVINSI," 2018. [Daring]. Tersedia pada: <https://www.depkes.go.id/>.
- [5] M. Arisanti dan N. H. Suryaningtyas, "KEJADIAN DEMAM BERDARAH DENGUE (DBD) DI INDONESIA TAHUN 2010-2019," *SPIRAKEL*, vol. 13, no. 1, hlm. 34–41, Des 2021, doi: 10.22435/spirakel.v13i1.5439.
- [6] R. A. Indraputra dan R. Fitriana, "K-Means Clustering Data COVID-19," *Jurnal Teknik Industri*, vol. 10, 2020.
- [7] R. R. Putra dan C. Wadisman, "Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 1, no. 1, hlm. 72–77, Mar 2018, doi: 10.31539/intecom.v1i1.141.
- [8] S. Parsaoran Tamba dan F. Toknady Kesuma, "PENERAPAN DATA MINING UNTUK MENENTUKAN PENJUALAN SPAREPART TOYOTA DENGAN METODE K-MEANS CLUSTERING," *Jurnal Sistem Informasi Ilmu Komputer Prima (JUSIKOM PRIMA)*, vol. 2, no. 2, 2019.
- [9] I. Nyoman dan M. Adiputra, "CLUSTERING PENYAKIT DBD PADA RUMAH SAKIT DHARMA KERTI MENGGUNAKAN ALGORITMA K-MEANS," *INSERT: Information System and Emerging Technology Journal*, vol. 2, no. 2, hlm. 99, 2021.
- [10] Z. Nabila, A. Rahman Isnain, dan 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, hlm. 100, 2021, [Daring]. Tersedia pada: <http://jim.teknokrat.ac.id/index.php/JTISI>
- [11] B. Rahmat dkk., "IMPLEMETASI K-MEANS CLUSTERING PADA RAPIDMINER UNTUK ANALISIS DAERAH RAWAN KECELAKAAN," 2018.