



Penerapan *Data Mining* Menggunakan Algoritma *K-Means* Untuk Menentukan Tingkat Vaksinasi Pada Kecamatan Tambun Selatan

Anissa Aidha Zulyani^{1✉}, Agung Susilo Yuda Irawan², Asep Jamaludin³

Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang

Email: anissa.aidha19007@student.unsika.ac.id^{1✉}

Abstrak

Berdasarkan penduduk yang sudah wajib di vaksin, total pencapaian sasaran penyebaran vaksinasi COVID-19 belum memenuhi target yang ditentukan. Penelitian ini bertujuan melakukan analisis *k-means clustering* terhadap penyebaran vaksinasi COVID-19 di kecamatan Tambun Selatan dengan variabel Vaksin 1, Vaksin 2, dan Vaksin 3. Menggunakan metode KDD dalam proses algoritma *k-means* dan dalam pengolahan data dibantu oleh *tools* Rstudio. Hasil *clustering* terbaik dengan fungsi *scale* dengan nilai *variance* 77,9%, serta melakukan evaluasi dengan hasil *clustering* terbaik dengan fungsi *scale* dengan validitas *connectivity* = 0,9111, validitas *dunn* = 0.2393, dan validitas *silhouette* = 0.6205. Hasil *clustering* menggunakan K = 3 menghasilkan *cluster* 1 = 20 RW dengan kategori pemerataan sangat merata, *cluster* 2 = 45 RW dengan kategori pemerataan sedang, *cluster* 3 = 115 RW dengan kategori pemerataan tidak merata.

Kata Kunci : *Vaksin COVID-19, Data Mining, KDD, K-Means, Clustering*

Abstract

Based on the population that has been compulsorily vaccinated, the total achievement of the COVID-19 vaccination distribution target has not met the specified target. This research is about the application of data mining using the k-means algorithm to determine the distribution level of COVID-19 vaccination. The aim is to carry out a k-means clustering analysis of the spread of COVID-19 vaccination in Tambun Selatan sub-district with Vaccine 1, Vaccine 2, and Vaccine 3 variables. Using the KDD method in processing the k-means algorithm and in data processing assisted by Rstudio tools. The best clustering results with the scale function with a variance value of 77.9%, as well as evaluating the best clustering results with the scale function with a connectivity validity = 0,9111, a dunn validity = 0.2393, and a silhouette validity = 0.6205. The results of clustering using K = 3 produce cluster 1 = 20 RW with very even distribution, cluster 2 = 45 RW with medium equity, cluster 3 = 115 RW with uneven distribution.

Keywords: *COVID-19 Vaccine, Data Mining, KDD, K-Means, Clustering*

PENDAHULUAN

Coronavirus atau *SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus2)* atau dikenal dengan COVID-19. COVID-19 adalah salah satu jenis virus *zoonosis* yang menyerang saluran pernafasan. Virus ini pertama kali ditemukan pada akhir tahun 2019 di kota Wuhan- China, lalu menjadi perhatian masyarakat dunia setelah pada 20 Januari 2020 yang kemudian menyebar dengan cepat keseluruh dunia termasuk negara Indonesia pada 2 Maret 2020.

Pemerintah telah menghimbau masyarakat untuk mendaftarkan diri dalam program vaksinasi dan terus menghimbau masyarakat untuk menjaga pola hidup yang sehat, mencuci tangan, dan menjaga jarak, serta menggunakan masker ketika berpergian keluar rumah serta untuk melawan virus COVID-19. Target tenaga kesehatan yang akan divaksinasi COVID-19 mencapai 1.468.764 orang. Sedangkan tahap dua untuk petugas pelayanan publik, lansia diatas 60 tahun hingga pedagang. pasar dilaksanakan sejak 17 Februari 2021. Target sasaran vaksinasi COVID-19 tahap dua mencapai 38.513.446 orang. Sedangkan vaksinasi tahap tiga merupakan masyarakat rentan terinfeksi COVID-19. Total pencapaian sasaran vaksinasi COVID-19 di Indonesia sebanyak 208.265.720 orang, dengan rincian vaksinasi dosis ke-1 sebanyak 201.379.534 orang, vaksinasi dosis ke-2 sebanyak 168.860.843 dan vaksinasi dosis ke-3 sebanyak 50.132.890 (M. Arief Utsman, 2022).

Total pencapaian sasaran penyebaran vaksinasi dari tahun 2019 sampai 2022 di kecamatan Tambun Selatan sebanyak 534.115 dosis vaksin, sedangkan target sasaran penyebaran vaksinasi COVID-19 dari tahun 2019 sampai 2022 di kecamatan Tambun Selatan

sebanyak 702,458 dosis vaksin, dengan kata lain target yang ditentukan belum lah memenuhi sasaran dalam penyebaran vaksin COVID-19, karena luasnya wilayah suatu daerah memungkinkan diperlukannya pengelompokkan bagian-bagian berdasarkan wilayah yang lebih kecil di daerah tersebut, pengelompokkan ini dimaksudkan untuk menghasilkan titik-titik pusat penyebaran kasus covid-19.

Penelitian terdahulu yang diselesaikan oleh Sindi et al (2022) dalam risetnya yang berjudul "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 di Indonesia". Karena belum adanya penelitian tentang penyebaran vaksinasi dan masih banyak masyarakat yang rentan akan penularan COVID-19 jadi dilakukan penelitian ini untuk mempermudah pekerjaan pihak dinas kesehatan kecamatan Tambun Selatan dalam masalah pengelompokkan vaksin dari setiap kelurahan yang ada di kecamatan Tambun Selatan (C. P. Ananda and E. Paujiah, 2021).

METODE PENELITIAN

Menggunakan metode data mining *Knowledge Discovery in Database* (KDD) dengan Algoritma K-means clustering yang kemudian dievaluasi menggunakan cluster validation seperti index validitas silhouette, index validitas connectivity dan index validitas dunn sehingga menghasilkan sebuah knowledge atau informasi.

Pengertian Data Mining menurut (S. S. M. K. Muhammad Arhami dan S. T. M. T. Muhammad Nasir, 2020) Data mining merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. Data mining mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data. Data mining juga dikenal sebagai Knowledge discovery, Knowledge extraction, data/pattern analysis, information harvesting, dan lain-lain.

Menurut (S. Aulia, 2021) Algoritma K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lainnya. Pengertian Validasi cluster merupakan suatu tahap kegiatan yang digunakan untuk memberikan penilaian hasil dari analisis suatu cluster dengan berdasarkan jumlah atau kuantitatif dan sesuai dengan keadaan yang sebenarnya atau objektif (D. sarjon Juliansa hengki dan Sumijan, 2017). Indeks validitas yang digunakan yaitu silhouette, dunn, dan connectivity. Indeks Silhouette merupakan metode yang dikembangkan oleh Rosseeuw pada tahun 1987. Indeks Silhouette menghasilkan cluster terbaik jika nilai yang diperoleh semakin dekat

dengan angka 1, rentang nilai serta interpretasi pada indeks silhouette dapat dilihat pada tabel dibawah.

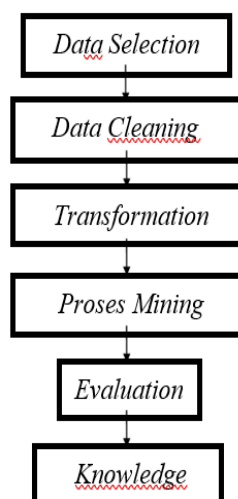
Tabel 2. 1 Rentang Nilai Silhouette dan Interpretasi

Rentang nilai	Interpretasi
0.71 - 1.0	Struktur yang kuat
0.51 - 0.70	Struktur yang beralasan
0.26 - 0.50	Struktur lemah
< 0.25	Tidak ditemukan struktur yang subtansial

(Sumber: D. Juliansa Hengki, Sumijan, 2017)

Sedangkan Indeks validitas Dunn (A. F. Khairati, A. . Adlina, G. . Hertono, dan B. . Handari, 2019) menghitung nilai minimum dari perbandingan antara nilai fungsi dissimilaritas antara dua klaster sebagai separation dan nilai maksimum dari diameter klaster sebagai compactness. Jumlah klaster terbaik ditunjukkan dengan semakin besar nilai DN. Dan Indeks Connectivity membentuk jumlah cluster terbaik apabila nilai yang dihasilkan semakin kecil dibandingkan dengan nilai dari cluster yang lain D. sarjon Juliansa hengki dan Sumijan, 2017).

Dalam menyelesaikan penelitian ini, diperlukan metodologi penelitian sebagai prosedur penelitian agar dalam penelitiannya berjalan dengan baik. Metodologi penelitian yang digunakanadalah metodologi Knowledge in Database (KDD) pada gambar 3.1.



Gambar 3.1 Metodologi Penelitian KDD

1. Data Selection

Data selection adalah proses seleksi data yang sesuai untuk dianalisis yang diambil dari sumber data atau database, seleksi data dilakukan karena tidak semua data yang ada pada database digunakan. Data yang diambil merupakan data atribut yang digunakan dosis vaksin 1, vaksin 2, dan vaksin 3 pada rentang waktu 2019 – 2022 dengan memiliki 181 baris dan 4 kolom dengan keterangan atribut awal NamaDesa_RW, Vaksin_1, Vaksin_2 dan Vaksin_3.

2. Pre-processing (Cleaning)

Data cleaning merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan.

3. Transformation

Transformasi data yaitu mengubah atau menggabungkan data ke dalam format yang sesuai untuk diproses dalam data mining. dengan skenario pertama data menggunakan fungsi scale dan skenario kedua data tidak menggunakan fungsi scale. Dalam proses transformasi data dilakukan pengubahan data dengan standarisasi data menggunakan fungsi scalling.

4. Proses Data Mining

Proses data mining merupakan proses mencari pengetahuan berharga dan tersembunyi dari data saat metode atau algoritma diterapkan dengan menentukan centroid secara acak.

5. Evaluation

Evaluasi digunakan untuk menerjemahkan pola dari hasil proses mining atau mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Serta dilakukan evaluasi terhadap hasil dari proses data mining tersebut menggunakan indeks validitas silhouette, dunn, dan connectivity dengan menentukan jumlah cluster.

6. Knowledge

Menyajikan pengetahuan atau memvisualisasikan hasil penelitian dengan bantuan tools aplikasi R.Studio menggunakan bahasa R. Visualisasi plot data ini dapat membantu untuk berbagi hasil penelitian data mining kepada pihak lain agar lebih mudah dipahami.

HASIL DAN PEMBAHASAN

Data Selection dari dataset memiliki 181 baris dan 4 kolom dengan keterangan atribut awal NamaDesa_RW, Vaksin_1, Vaksin_2 dan Vaksin_3, serta Atribut yang digunakan adalah Vaksin Dosis 1, Vaksin Dosis 2, dan Vaksin Dosis 3. Sampel data tabel.

Tabel 4.1 Dataset Vaksinasi COVID-19

Nama_Desa	Vaksin_1	Vaksin_2	Vaksin_3
Jatimulya	43752	40832	7743
Lambang Sari	12354	11960	2024
Lambangjaya	1629	1119	472
Mangunjaya	44819	41588	12508
Mekarsari	24050	22254	5398
Setiadarma	4772	4320	2734
Setiamekar	32451	30125	5395
Sumberjaya	46902	39244	15140
Tambun	17267	14798	6752
Tridayasakti	18655	17150	5908

Setelah dilakukan data cleaning hasil summary dari dataset yang digunakan, pada hasil summary tersebut tidak ditemukan nilai NA atau missing value.

NamaDesa_RW	Vaksin_1	Vaksin_2	Vaksin_3
Length:180	Min. : 162.0	Min. : 115	Min. : 39.0
Class :character	1st Qu.: 741.8	1st Qu.: 685	1st Qu.: 132.8
Mode :character	Median : 995.0	Median : 919	Median : 249.5
	Mean :1370.3	Mean :1241	Mean : 356.0
	3rd Qu.:1893.8	3rd Qu.:1759	3rd Qu.: 475.0
	Max. :4374.0	Max. :3948	Max. :1757.0

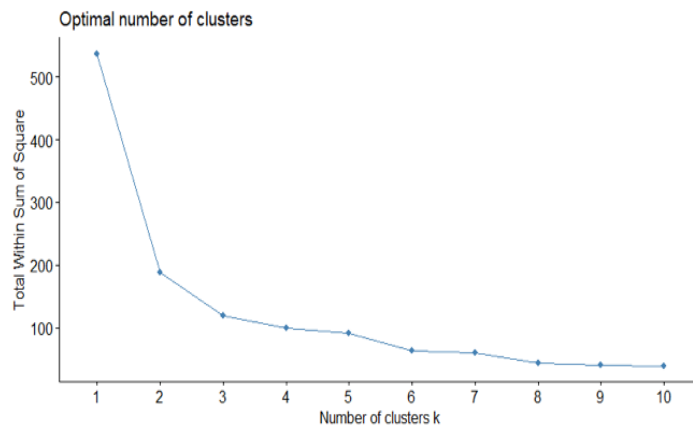
Gambar 4.1 Hasil Summary Dataset

Tahapan transformasi data dengan normalisasi menggunakan fungsi scale atau skala untuk semua nilai data, agar beberapa variabel yang digunakan memiliki rentang nilai yang sama dan tidak ada yang terlalu besar atau terlalu kecil. Sampel data tabel.

Tabel 4.2 Hasil Scalling

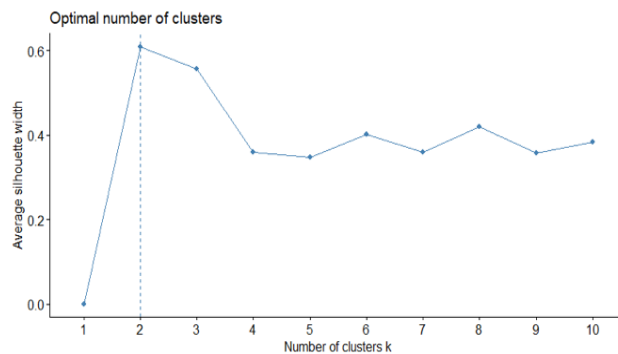
NamaDesa _RW	Vaksin_1	Vaksin_2	Vaksin_3
Jatimulya	2,09739193 9	2,4299728 5	1,5489317
Lambangsa ri	- 0,90776321 2	-0,87717	- 0,9565042
Lambangja ya	- 1,321960217	- 1,3778382	- 1,0265863
Mangunjay a	0,44283078 5	0,3043073 9	0,4416341 7
Mekarsari	- 0,65278709 9	- 0,6547889	- 0,7742906
Setiadarma	- 0,941166196	-0,936803	- 0,7742906
Setiamekar	0,41276809 9	0,3937567 8	- 0,6446387
Sumberjaya	0,22459795 5	0,2235544 7	2,2217200 7
Tambun	1,74888747	1,6361094 8	1,8082355 5
Tridayasakt i	- 0,70066470 9	- 0,6957865	- 0,2206419

Pada proses data mining menentukan jumlah cluster terbaik atau menentukan nilai K optimal dengan beberapa metode seperti elbow method, silhouette method, dan diagram gap statistic.



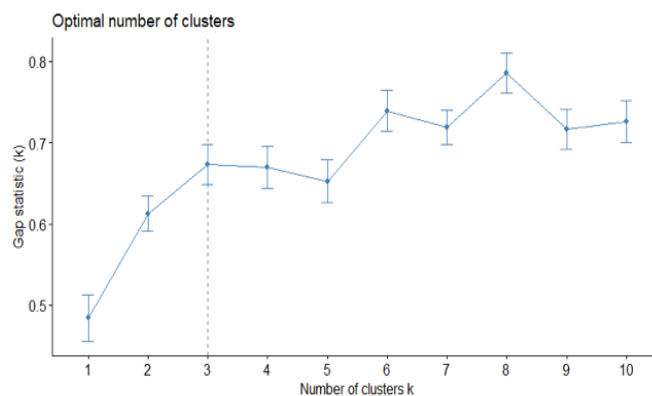
Gambar 4. 2 Grafik Elbow Method

Metode elbow baik pada data dengan penskalaan maupun tanpa penskalaan sama-sama memiliki cluster optimal sejumlah 3 cluster.



Gambar 4. 1 Grafik *Silhouette Method*

Sedangkan grafik silhouette method untuk penentuan nilai K optimum atau terbaik dilihat dari garis yang dihubungkan dengan garis putus-putus, dengan kata lain grafik silhouette method menunjukkan nilai K terbaik adalah K=2.



Gambar 4. 4 Grafik Gap Statistic

Dan metode GAP Statistic / gap_stat dilihat dari garis yang dihubungkan dengan garis putus-putus, dengan kata lain yang menunjukkan bahwa nilai K terbaik adalah saat K=3.

Proses evaluasi dengan 2 skenario yaitu menggunakan fungsi scaling dan tanpa menggunakan fungsi scaling dimana menggunakan penentuan 3 cluster dengan

menggunakan indeks validitas silhouette, dunn, dan connectivity.

```
Clustering Methods:
kmeans

Cluster sizes:
2 3

Validation Measures:
                2      3

kmeans Connectivity 13.4444 0.9111
      Dunn          0.0598 0.2393
      Silhouette    0.6205 0.6054

Optimal Scores:

      Score Method Clusters
Connectivity 0.9111 kmeans 3
      Dunn    0.2393 kmeans 3
      Silhouette 0.6205 kmeans 2
```

Gambar 4. 2 Hasil *Evaluation* Fungsi *Scalling*

```
Clustering Methods:
kmeans

Cluster sizes:
2 3

Validation Measures:
                2      3

kmeans Connectivity 7.2302 18.5480
      Dunn          0.0639 0.0814
      Silhouette    0.6618 0.6203

Optimal Scores:

      Score Method Clusters
Connectivity 7.2302 kmeans 2
      Dunn    0.0814 kmeans 3
      Silhouette 0.6618 kmeans 2
```

Gambar 4. 3 Hasil *Evaluation* Tanpa *Scalling*

Hasil perbandingan evaluasi dengan percobaan menggunakan penentuan 3 cluster dengan indeks validitas clustering menggunakan fungsi scale dan tanpa menggunakan fungsi scale, Nilai indeks connectivity terbaik adalah dengan menggunakan fungsi scale, nilai indeks dunn terbaik adalah dengan menggunakan fungsi scale, sedangkan pada nilai indeks silhouette dengan nilai struktur yang kuat adalah tanpa menggunakan fungsi scale.

Menyajikan data agar lebih mudah untuk dipahami dan dibaca. Dalam tahap penyajian dari hasil penelitian yang telah dilakukan, menunjukkan hasil akhir clustering dengan penggunaan 3 cluster agar menyajikan data lebih mudah untuk dipahami dan dibaca. Pada tahapan data mining hasil clustering hanya berupa informasi vector dan visualisasi plot. Data pada tabel 4.7 merupakan sebuah sampel. Detail hasil final Clustering K=3 terlihat pada Lampiran 5, dan hasil visualisasinya terlihat pada Gambar 4.9.

Tabel 4.1 Hasil *Final Clustering* K=3

NamaDesa_RW	Cluster_Final
Jatimulya_1	1
Lambang Sari_1	3
Lambangjaya_1	3
Mangunjaya_1	2

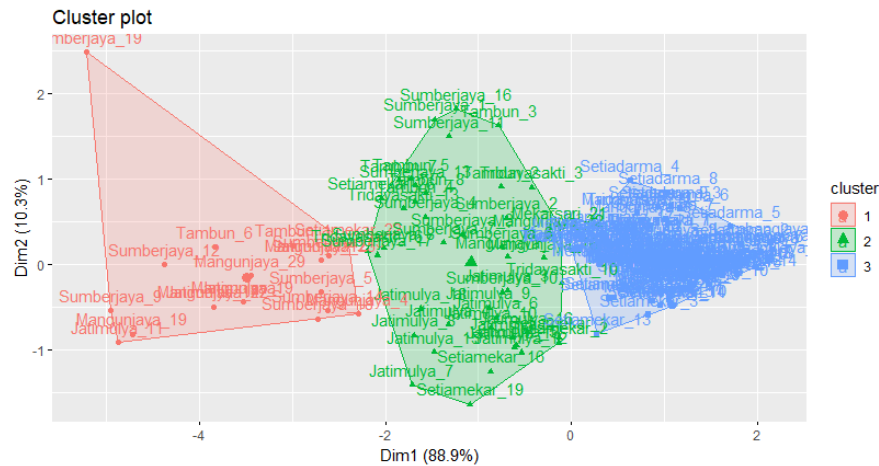
NamaDesa_RW	Cluster_Final
Mekarsari_1	3
Setiadarma_1	3
Setiamekar_1	2
Sumberjaya_1	2
Tambun_1	1
Tridayasakti_1	3

Tahapan knowledge hasil clustering berupa informasi vector dan visualisasi plot.
sampil tabel hasil final Clustering K=3 dengan kategori.

Tabel 4.2 Hasil *Clustering* dengan Kategori

Cluster	PerDesa_RW	Karakteristik			Jumlah	Kategori
		Vaksin_1	Vaksin_2	Vaksin_3		
C1	Jatimulya_1, 11, 14, 17	1,236708378 s/d	0,86460846 s/d	0,75700400 s/d	20	Sangat Merata
	Mangunjaya_4, 9, 19, 22, 25, 29	3,34443669	3,36297973	4,90936945		
	Setiamekar_22					
	Sumberjaya_5, 9, 12, 14, 15, 18, 19					
	Tambun_1 dan 6					
C2	Jatimulya_2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 15, 16, 18	-0,16510353 s/d 1,71993822	- 0,35165483 s/d 1,65350242	- 0,78129890 s/d 2,22172007	45	Pemerataan Sedang
	Mangunjaya_1, 16, 32					
	Mekarsari_21					
	Setiamekar_1, 2, 16, 19, 26					
	Sumberjaya_1, 2, 3, 4, 6, 7, 8, 10, 11, 13, 16, 17					
	Tambun_2, 3, 4, 5, 7, 8					

Cluster	PerDesa_RW	Karakteristik			Jumlah	Kategori
		Vaksin_1	Vaksin_2	Vaksin_3		
	Tridayasakti_3, 10, 13, 16					
C3	Lambang Sari_1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21	-1,34534231 s/d 0,12216214	- 1,39895820 s/d 0,16764859	-1,11068480 s/d 0,47317113	115	Tidak Merata
	Lambangjaya_1, 2, 3, 4, 5, 6					
	Mangunjaya_3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 17, 18, 20, 21, 23, 24, 26, 27, 28, 30, 31					
	Mekarsari_1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25					
	Setiadarma_1, 2, 3, 4, 5, 6, 7, 8, 9,					
	Setiamekar_3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 20, 21, 23, 24, 25					
	Tridayasakti_1,2, 4, 5, 6, 7, 8, 9, 11, 12, 14, 15					



Gambar 4. 4 Hasil Plot Visualisasi *Clustering* K=3

SIMPULAN

Penerapan teknik clustering menggunakan algoritme k-means pada data vaksin di Kec. Tambun Selatan pada tahun 2019 - 2022 dengan tools Rstudio menghasilkan jumlah cluster optimal yaitu K=3. Pada pemerataan daerah berdasarkan 180 RW di Kec. Tambun Selatan sebanyak 20 RW masuk kedalam kategori sangat merata pada cluster 1, 45 RW masuk kedalam kategori pemerataan sedang pada cluster 2, 115 RW masuk kedalam kategori tidak merata pada cluster 3.

DAFTAR PUSTAKA

- C. P. Ananda and E. Paujiah, "Sosialisasi Vaksinasi Covid-19 Melalui Media Cetak untuk Meningkatkan Pemahaman Masyarakat Socialization of the Covid-19 Vaccination Through Print Media to Improve Public Understanding About the Importance of the Covid-19 Vaccination," vol. 32, no. November, 2021.
- D. P. Utomo and B. Purba, "Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 846, 2019, doi: 10.30645/senaris.v1i0.91.
- D. sarjon Juliansa hengki and Sumijan, "Uji Validasi Algoritma Self Organizing Map (SOM) dan K-Mens untuk Pengelompokan Pegawai," *Resti*, vol. 1, no. 1, pp. 19–25, 2017.
- F. Rahmawati and N. Merlina, "Metode Data Mining Terhadap Data Penjualan Sparepart Mesin Fotocopy Menggunakan Algoritma Apriori," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 6, no. 1, pp. 9–20, 2018, doi: 10.33558/piksel.v6i1.1390.
- H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, p. 62, 2019, doi: 10.26418/jp.v5i1.29611.

- J. Schmuller, *Statistical Analysis with R For Dummies*. Wiley, 2017. [Online]. Available: <https://books.google.co.id/books?id=HtlGDgAAQBAJ>
- K. Erwansyah, "Implementasi Data Mining Untuk Menganalisa Hubungan Data Penjualan Produk Bahan Kimia Terhadap Persediaan Stok Barang Menggunakan Algoritma FP (Frequent Pattern) Growth Pada PT . Grand Multi Chemicals," *J. Teknol. Sist. Inf. dan Sist. Komput. TGD (J-SISKO TECH)*, vol. 2, no. 2, pp. 30–40, 2019.
- M. E. Lestiani, "Faktor-Faktor Dominan Promosi Yang Mempengaruhi Motivasi Konsumen Dalam Membeli Suatu Produk Dengan Menggunakan Metode Ahp," *Ind. Elektro Penerbangan*, vol. 1, no. Vol. 1 No. 1 (2011): INDEPT, p. 20, 2020, [Online]. Available: https://www.google.co.id/url?sa=t&rct=j&q=&esrc=s&source=web&cd=26&cad=rja&uact=8&ved=2ahUKEwj9nKqAws_eAhWNXisKHYPcyl4FBAWMAV6BAgHEAI&url=http%3A%2F%2Fjip.polinema.ac.id%2Fjojs3%2Findex.php%2Fjip%2Farticle%2Fview%2F80%2F79&usg=AOvVaw2sufCT1LINuHSa_Bu4PkQL
- M. R. D. Jodi, "Algoritma dan Struktur Data," *Fak. Komput.*, vol. 2, no. 1, pp. 1–10, 2020.
- M. Wahyudi, M. Masitha, R. Saragih, S. Solikhun, and J. Simarmata, *Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*. Yayasan Kita Menulis, 2020. [Online]. Available: <https://books.google.co.id/books?id=wQnhDwAAQBAJ>
- N. L. W. S. R. Ginantra *et al.*, *Data Mining dan Penerapan Algoritma*. Yayasan Kita Menulis, 2021. [Online]. Available: <https://books.google.co.id/books?id=v0gtEAAAQBAJ>
- P. Zimmermann and N. Curtis, "Coronavirus infections in children including COVID-19: An overview of the epidemiology, clinical features, diagnosis, treatment and prevention options in children," *Pediatr. Infect. Dis. J.*, vol. 39, no. 5, pp. 355–368, 2020, doi: 10.1097/INF.0000000000002660.
- S. S. M. K. Muhammad Arhami and S. T. M. T. Muhammad Nasir, *Data Mining - Algoritma dan Implementasi*. Penerbit Andi, 2020. [Online]. Available: <https://books.google.co.id/books?id=AtcCEAAAQBAJ>
- R. Hermawan, "Analisa data mining untuk segmentasi daerah prioritas penerima vaksin covid-19 menggunakan algoritma k- means pada provinsi jawa barat".ss
- R. M. Abarca, "Analisis Cluster Sebaran Covid-19 Menggunakan Algoritma K-Means Clustering (Studi Kasus: Provinsi Jawa Barat) Cluster," *J. Inf. Technol. Comput. Sci. Vol. 4 Nomor 2, Desember 2021 e-ISSN 2614-1574 p-ISSN 2621-3249 Anal.*, vol. 4, no. 2020, pp. 2013–2015, 2021.
- R. Muliono and Z. Sembiring, "Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 4, no. 2, pp. 2502–714, 2019.
- S. Aulia, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering

- (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, pp. 1–5, 2021, doi: 10.46576/djtechno.v1i1.964.
- Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- Z. Abidin, A. Hudaya, and D. Anjani, "Efektivitas Pembelajaran Jarak Jauh Pada Masa Pandemi Covid-19," *Res. Dev. J. Educ.*, vol. 1, no. 1, p. 131, 2020, doi: 10.30998/rdje.v1i1.7659.
- Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>