



## *Application of Data Mining to Covid-19 Vaccination Based on Side Effects After Vaccination Using the K-Medoids Clustering Algorithm*

### **Penerapan Data Mining Terhadap Vaksinasi Covid-19 Berdasarkan Efek Samping Setelah Vaksin Menggunakan Algoritma K-Medoids Clustering**

**Alji Ridwan Syah Alam<sup>1</sup>, Agisti Mutiara Ayulya<sup>2</sup>, Agus Syaifullah<sup>3</sup>,  
Della Oktoriani<sup>4</sup>, Mutiara Rahmadani<sup>5</sup>**

<sup>1,2,3,4,5</sup>Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,  
UIN Sultan Syarif Kasim Riau, Indonesia

E-Mail: <sup>1</sup>12050311668@students.uin-suska.ac.id, <sup>2</sup>12050322046@students.uin-suska.ac.id,  
<sup>3</sup>12050317454@students.uin-suska.ac.id, <sup>4</sup>12050326263@students.uin-suska.ac.id,  
<sup>5</sup>12050321570@students.uin-suska.ac.id

*Corresponding Author: Alji Ridwan Syah Alam*

#### **Abstract**

Data mining is a process of collecting and processing data that aims to produce important information on data. Data mining can be applied in various fields such as the Covid-19 vaccination. Vaccination aims to actively increase a person's immunity to prevent getting Covid-19. Now vaccines have been developed and have different side effects. The amount of information that comes in from various sources, can affect people's thinking, thus getting doubts and rejections from the public about this vaccination. The data were analyzed using the K-Medoids algorithm method. This method identifies the same groups and traits based on certain criteria. The results of research with the K-Medoids algorithm looked at the side effects felt by the community after vaccination. This study aims to determine the side effects felt by the community after vaccination and group them by vaccine type, age, and side effect impact. The impact of side effects is divided into 3, namely: Non-effects, Weight, Light. Research is carried out based on finding knowledge by providing data in the form of lift with an online survey design. The online survey was conducted using purposive sampling techniques by filling out questionnaires to the public.

*Keyword: Clustering, Covid-19, K-Medoids, Side Effects, Vaccines*

#### **Abstrak**

Data mining merupakan proses pengumpulan dan pengolahan data yang bertujuan untuk menghasilkan informasi penting pada data. Data mining bisa di terapkan dalam berbagai bidang seperti halnya yaitu vaksinasi Covid-19. Vaksinasi bertujuan untuk meningkatkan kekebalan tubuh seseorang secara aktif agar mencegah terkena Covid-19. Sekarang vaksin sudah di kembangkan dan memiliki efek samping yang berbeda-beda. Banyaknya informasi yang masuk dari berbagai sumber, dapat mempengaruhi pemikiran masyarakat, sehingga medapat keraguan dan penolakan dari masyarakat akan dari vaksinasi ini. Data di analisa menggunakan metode algoritma K-Medoids. Metode ini mengidentifikasi kelompok dan ciri-ciri yang sama berdasarkan kriteria tertentu. Hasil penelitian dengan algoritma K-Medoids melihat efek samping yang di rasakan masyarakat pasca Vaksinasi. Penelitian ini bertujuan untuk mengetahui efek samping yang di rasain oleh masyarkat pasca vaksinasi dan mengelompokkannya berdasarkan jenis vaksin, usia, dan dampak efek samping. Dampak efek samping di bagi menjadi 3 yaitu : Non-efek, Berat, Ringan. Penelitian di lakukan berdasarkan menemukan pengetahuan dengan memberi data berupa angkat dengan desain *survey online*. *Survey online* dilakukan menggunakan teknik purposive sampling dengan mengisi kuisioner terhadap masyarakat.

*Kata Kunci: Clustering, Covid-19, Efek Samping, K-Medoids, Vaksin*

## 1. PENDAHULUAN

Virus Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) atau bisa disebut virus Corona adalah virus yang menyerang sistem pernafasan, pneumonia akut, sampai kematian. Penyakit yang diakibatkan oleh virus ini disebut COVID-19 [1]. Covid-19 adalah penyakit yang belum pernah diketahui oleh manusia [2]. Virus ini dapat menyerang siapa saja dan di mana saja, biasanya virus ini menyerang orang dewasa, ibu yang sedang hamil, anak-anak, dan lain-lain [3]. Penyebaran covid terjadi dengan sangat cepat, semua negara di dunia memberitahukan kepada penduduknya akan covid-19 [4].

Virus ini berasal dari negara China tepatnya di kota Wuhan pada tahun 2019, kemudian ditetapkan oleh WHO sebagai wabah pandemi pada tanggal 12 Maret 2020 [5]. Gejala yang timbul tergantung pada jenis virus yang menyerang. gejala dapat dirasakan setelah 5-6 hari sejak terpapar virus ini atau selambatlambatnya 14 hari [6].

Untuk memutus mata rantai penularan virus ini, yaitu dengan menerapkan protokol Kesehatan (menerapkan 5M), sementara itu vaksin menjadi salah satu solusi terbaik yang sedang dikembangkan untuk mengendalikan covid-19 [7]. World Health Organisation (WHO) menginformasikan berbagai jenis vaksin diantaranya yaitu vaksin Pfizer, vaksin Sinovac, vaksin Moderna, dan vaksin Astrazeneca, [8]. Sehingga terselenggaranya vaksinasi massal di Indonesia yang dijalani hampir seluruh masyarakat dengan dosis 1 mampu memberikan dampak yang besar agar masyarakat lebih memahami pentingnya melakukan vaksinasi tersebut [9]. Program vaksinasi massal ini mendapat banyak hambatan dari masyarakat, karena adanya efek samping yang diterima oleh penerima vaksin covid-19 [10] [11]. Dikarenakan banyaknya hambatan dari masyarakat inilah maka diperlukan pengelompokan berdasarkan efek samping dari vaksinasi. Pengelompokan atau menganalisis data dalam jumlah yang berskala besar dibutuhkan sebuah algoritma data mining pengelompokan [12].

Data Mining adalah tahapan pencarian pola-pola yang tersembunyi yaitu ilmu yang belum diketahui sebelumnya dari suatu kelompok data yang berada di dalam data warehouse atau database. Data mining memiliki tugas dan yang dapat dilakukannya seperti, deskripsi dimana analisis menggambarkan pola dan metode di dalam data. Selanjutnya, Estimasi yang lebih mencondong kearah numerik dari pada ke arah kategori. Selain itu, tugas yang lainnya yaitu untuk memprediksi yang mana hasil dan nilai sudah diprediksi [13].

Clustering adalah suatu proses pengelompokan pada beberapa kelompok data. Pengelompokan Hal tersebut untuk membuktikan pola yang serupa pada tipe cluster yang sejenis. Hasil dari metode ini adalah membuat pengelompokan yang sama dengan bentuk grup. Tujuannya adalah untuk mengidentifikasi perbedaan antara kategori [14].

Berdasarkan pada penelitian sebelumnya oleh S. Sindi, dkk, menyebutkan bahwa melakukan pengelompokan untuk masyarakat yang terjangkit covid-19 menggunakan metode K-Medoids [15]. Sedangkan menurut B. Purba dan E. Bu'ulolo menjelaskan klaster covid di Sumatra Utara dan hasilnya membagi zona daerah di Sumatra Utara menjadi tiga zona yaitu zona merah, kuning, dan hijau didasari oleh jumlah kasus covid-19 dan kematian yang di sebabkan oleh covid-19 [16].

Dari paparan di atas mendorong penulis untuk melakukan penelitian dengan judul "Penerapan Data Mining Untuk Mengetahui Efek Samping Setelah Vaksinasi Covid-19 Menggunakan Algoritma K-Medoids Clustering".

Pada penelitian ini juga penulis melakukan survey online melalui pengisian quisioner efek samping setelah vaksinasi Covid-19 yang dikelompokkan dalam 3 kategori yaitu tidak merasakan efek samping (non effect), ringan, hingga berat. Didalam pengelompokan data efek samping tersebut, menggunakan metode K-Medoids Clustering, metode ini termasuk dalam varian algoritma K-Means, yang bersifat efisien dan digunakan untuk tipe data berkategori. Data yang didapatkan dari penelitian ini diharapkan bisa memberikan informasi dan menghilangkan keraguan masyarakat untuk melakukan vaksinasi.

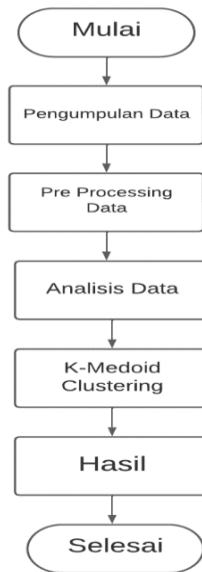
## 2. METODOLOGI PENELITIAN

### 2.1 Data Mining

Data mining, yang dikenal sebagai *Knowledge Discovery in Database (KDD)* dalam database, melibatkan penggunaan data, pengumpulan, dan data historis untuk menemukan pola dan hubungan kumpulan data besar pola data secara otomatis terdeteksi. Setiap sampel yang diperoleh harus memberikan keuntungan dan manfaat tertentu [17]. Pemilihan algoritma atau metode yang tepat sangat bergantung untuk tujuan dan proses KDD secara keseluruhan [18].

### 2.2 Tahapan Penelitian

Penelitian ini dilakukan secara kuantitatif dengan *desain survey cross sectional*. *Desain survey cross sectional* merupakan survei yang dikaitkan dengan pengisian quisioner, pengumpulan dilakukan dalam waktu tertentu dengan jumlah data lebih dari satu variabel. Data yang dikumpulkan berupa data kuantitatif maupun kualitatif.



**Gambar 1.** Metodologi Penelitian

Penelitian ini mengimplementasikan beberapa tahapan utama seperti yang tertera pada gambar 1, diantaranya pengumpulan data, *pre processing* data, proses analisis data, proses klasterisasi menggunakan algoritma K-Medoids serta yang terakhir tahap hasil.

### 2.3 Clustering

Clustering adalah proses pengelompokan dataset, observasi, atau kelas yang memiliki objek yang sama. [19]. Clustering telah menjadi instrument valid dalam memecahkan masalah -masalah kompleks ilmu komputer dan statistika [20]. Mengelompokkan atau mengkategorikan objek berdasarkan informasi yang dikumpulkan dari data yang menggambarkan hubungan antar objek, berdasarkan prinsip memaksimalkan kesamaan antar anggota kelas dan meminimalkan kesamaan antar kelas/cluster [21].

### 2.4 K-Medoids

(*PAM*) *Partitioning Around Medoids* atau K-Medoids adalah metode pengelompokan data yang masih berkaitan dengan algoritma K-Means. Algoritma ini bertujuan untuk meminimalisir kerentanan partisi ekstrim pada record data [22] [14]. Masing-masing data dibagikan ke pengelompokan yang terdekat dengan persamaan Euclidia.

$$dik \sqrt{\sum_{j=1}^m (c_{ij} - x_{ik})} \quad (1)$$

### 2.5 Rapid Miner

Rapidminer adalah perangkat lunak manajemen penambangan data. Aktivitas Rapidminer Text Mining adalah mengekstrak pola dari kumpulan data besar dan menggabungkannya dengan teknik statistik seperti analisis teks, kecerdasan buatan, dan basis data. [23].

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan dan Pre-Processing Data

Data yang digunakan dalam penelitian ini diperoleh dari respon masyarakat melalui pengisian *quisioner* berdasarkan *survey online* dengan menggunakan teknik *purposive sampling*. Terdapat 118 data yang diperoleh dari respon masyarakat terhadap vaksinasi covid-19 berdasarkan efek samping setelah vaksin, Data tersebut terdiri dari jenis vaksin *Sinovac* berjumlah 91 data, jenis vaksin *Astrazaneca* berjumlah 11 data, jenis vaksin *Pfizer* berjumlah 9 data, jenis vaksin *Morderna* berjumlah 5 data, dan jenis vaksin *Biofarma* berjumlah 1 data. Sementara itu diperoleh hasil efek samping dari beberapa jenis vaksin tersebut diantaranya terdapat nyeri, demam, batuk, sesak nafas, mengantuk, sakit dada dan tidak merasakan efek samping. Kemudian dilakukan proses transformasi data sehingga dapat diolah menggunakan algoritma K-Medoids Clustering.

**Table 1.** Data Survey Online

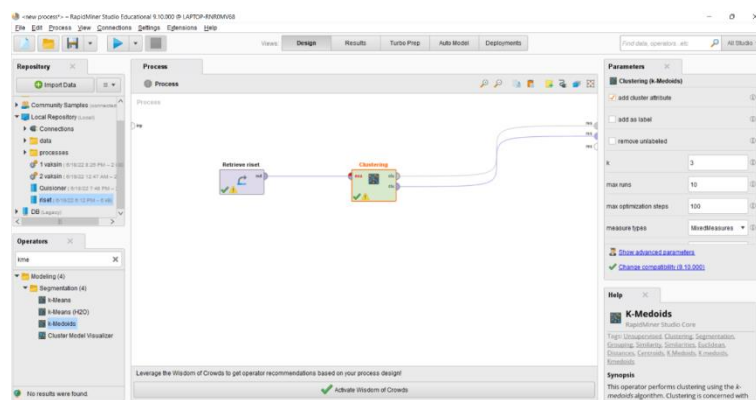
No	Nama	Jenis Kelamin	Umur	Jenis Vaksin Ke-1	Efek Samping Ke-1
1	NANDA TRY LUCHIA	Perempuan	20	Morderna	Nyeri,Batuk,Mengantuk
2	YULIANA PUTRI	Perempuan	19	Sinovac	Tidak Ada Efek
3	RACHELL APRINASTYA	Perempuan	19	Sinovac	Tidak Ada Efek
4	DIHAN AIBAGA	Laki-Laki	21	Pfizer	Tidak Ada Efek
5	IMAN SYAVIAN	Laki-Laki	19	Sinovac	Tidak Ada Efek
6	NAILUL AMANI	Perempuan	20	Sinovac	Tidak Ada Efek
7	M. FADLI	Laki-Laki	23	Sinovac	Nyeri,Batuk,Mengantuk
8	REZKY AMANDA	Perempuan	19	Sinovac	Tidak Ada Efek
9	DELVI NUR AINI	Perempuan	20	Sinovac	Tidak Ada Efek
10	FADLI GUSTAME	Laki-Laki	23	Sinovac	Nyeri,Batuk,Mengantuk
11	PIAN	Laki-Laki	23	Sinovac	Tidak Ada Efek
12	DIAN RIZKI	Laki-Laki	23	Sinovac	Demam,Sesak Nafas, Sakit dada
13	NAYA SYAHWITA	Perempuan	17	Sinovac	Tidak Ada Efek
14	ARDIKUN	Laki-Laki	22	Biofarma	Nyeri,Batuk,Mengantuk
15	ZAKIUL ADHA	Laki-Laki	20	Pfizer	Nyeri,Batuk,Mengantuk
16	IKHSAN FADLY	Laki-Laki	19	Sinovac	Tidak Ada Efek
17	MUCHLIS ALAM SYAH	Laki-Laki	20	Pfizer	Nyeri,Batuk,Mengantuk
18	LUKMAN	Laki-Laki	48	Sinovac	Tidak Ada Efek
19	DELLA HARMUTIKA	Perempuan	19	Astrazeneca	Nyeri,Batuk,Mengantuk
20	AULIA	Perempuan	20	Sinovac	Tidak Ada Efek
21	M.HAMDANI	Laki-Laki	18	Sinovac	Nyeri,Batuk,Mengantuk
22	RA	Laki-Laki	20	Sinovac	Nyeri,Batuk,Mengantuk
23	FENNY	Perempuan	20	Sinovac	Tidak Ada Efek
24	RAHMA SANI N.	Perempuan	19	Morderna	Tidak Ada Efek
25	FEBIOLA SISKAMUHAMMAD MUFTI	Perempuan	20	Sinovac	Tidak Ada Efek
26	MAHRI	Laki-Laki	20	Astrazeneca	Nyeri,Batuk,Mengantuk
27	FITO CAHYA K.	Laki-Laki	20	Pfizer	Tidak Ada Efek
...	...	...	...	...	...
118	DWI SEPTIANI	Perempuan	19	Sinovac	Tidak Ada Efek

Repon yang tidak sesuai dengan kategori seperti tidak mengisi jenis vaksin maka tidak diikutsertakan dalam tahap tranformasi data. Hal ini dilakukan untuk mencegah terjadinya ketidak sesuaian data Cluster yang dihasilkan.

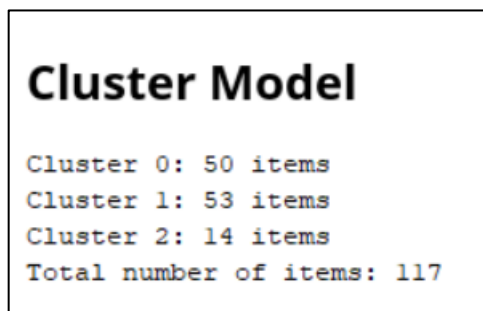
Selanjutnya dilakukan tahap normalisasi sebanyak 117 data yang berhasil lanjut dari tahap sebelumnya. Pada tahap normalisasi data menggunakan rumus *min-max normalization*. Normalisasi dilakukan untuk mengubah data agar nilai menjadi lebih mudah dipahami. Persiapan data ini merupakan salah satu tahapan merapikan respon masyarakat dan mengelompokkan berdasarkan variable yang ditentukan.

### 3.2 Pengelompokan dan Pengujian Data Menggunakan Algoritma K-Medoids

Pada tahapan ini dilakukan pengelompokan data Vaksinasi Covid-19. Langkah awal dilakukan dengan memasukkan data yang telah dinormalisasi ke dalam RapidMiner, Selanjutnya dilakukan tahap pengelompokan dengan menggunakan algoritma K-Medoids dan menentukan nilai K yang diinginkan, pada penelitian ini melakukan pengelompokan menjadi 3 Cluster.

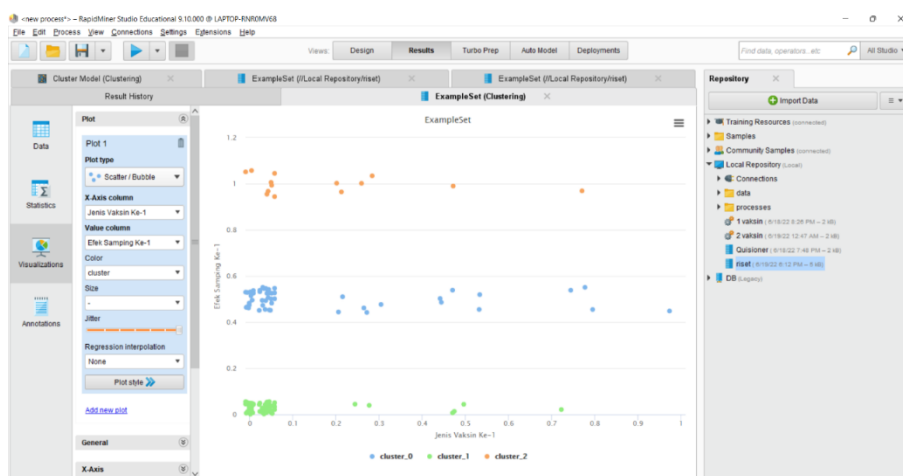
**Gambar 2.** Hasil Penerapan K-Medoids

Setelah melakukan desain, selanjutnya menjalankan project dan didapatkan hasil clustering setelah melakukan running.



Gambar 3. Hasil Clustering

Sehingga dapat diketahui persebaran Cluster melalui pola persebaran Cluster.



Gambar 4. Hasil Persebaran Cluster

#### Keterangan :

- Cluster 0 : Efek Pasca Vaksinasi Ringan
- Cluster 1 : Efek Pasca Vaksinasi No Effect
- Cluster 2 : Efek Pasca Vaksinasi Berat

Cluster pertama atau cluster 0 merupakan cluster dengan kategori ber-efek samping ringan yang berjumlah 50 data yang ditandai dengan nyeri, batuk, dan mengantuk. Pada cluster kedua atau cluster 1 didominasi oleh responden yang tidak memiliki efek samping sebanyak 53 data, selanjutnya cluster ketiga atau cluster 2 masuk dalam kategori berat yang berjumlah 14 data dengan ditandai adanya demam, sesak nafas, dan sakit dada setelah vaksinasi.

#### 4. KESIMPULAN

Berdasarkan hasil yang diperoleh dari pengelompokan efek samping pasca vaksinasi Covid-19, dihasilkan ada tiga clustering dengan karakteristik yang berbeda. Pada data clustering pertama mewakili respon masyarakat yang menerima efek samping ringan yang ditandai adanya nyeri, batuk, dan mengantuk. Data clustering kedua mewakili respon masyarakat yang tidak merasakan efek samping, serta data clustering ketiga ditandai adanya Demam, Sesak Nafas, dan Sakit dada mewakili efek samping dengan kategori berat. Dengan adanya data yang dihasilkan dapat menjadi acuan bagi masyarakat untuk melaksanakan program vaksin. Selain itu dapat dimanfaatkan pula untuk mencegah adanya informasi-informasi yang menyebabkan penolakan dari masyarakat untuk melaksanakan program vaksin.

#### REFERENSI

- [1] S. Timah, "Hubungan Penyuluhan kesehatan dengan Pencegahan covid 19 di Kelurahan kleak kecamatan Malalayang Kota Manado," *Indones. J. Community Dedication*, vol. 3, no. 1, pp. 7–14, 2021.

- [2] N. K. S. Astini, "Pemanfaatan teknologi informasi dalam pembelajaran tingkat sekolah dasar pada masa pandemi covid-19," *Lampuhyang*, vol. 11, no. 2, pp. 13–25, 2020.
- [3] A. Info and A. History, "Analisis Masalah Psikologis pada Ibu Hamil Selama Masa Pandemi Covid-19 : Literature Review," pp. 9–15, 2020.
- [4] N. Nurhalimah, "Upaya Bela Negara Melalui Sosial Distancing Dan Lockdwon Untuk Mengatasi Wabah Covid-19," *Sekol. Tinggi Tarb. Insa. Kamil*, pp. 1–6, 2020.
- [5] N. N. Kharisma, M. V. Roesminingsih, and S. Suhanadji, "Gambaran kebutuhan pembelajaran daring pkbm budi utama surabaya pada masa pandemi covid-19," *J. Pendidik. Nonform.*, vol. 15, no. 1, pp. 38–44, 2020.
- [6] R. Ayunda, V. Kosasih, and H. S. Disemadi, "Perlindungan hukum bagi masyarakat terhadap efek samping pasca pelaksanaan vaksinasi covid-19 di Indonesia," *Nusant. J. Ilmu Pengetah. Sos.*, vol. 8, no. 3, pp. 194–206, 2021.
- [7] A. R. Wijayanti, Y. D. Pora, O. Irman, Y. Nelista, and Y. M. H. Keytimu, "Psychological Conditions of Medical Personnel in Facing the Pandemic of Covid 19: Systematic Review," in *1st UMGESHIC International Seminar on Health, Social Science and Humanities (UMGESHIC-ISHSSH 2020)*, 2021, pp. 652–657.
- [8] S. A. Nugroho and I. N. Hidayat, "Efektivitas Dan Keamanan Vaksin Covid-19: Studi Refrensi," *J. Keperawatan Prof.*, vol. 9, no. 2, pp. 61–107, 2021.
- [9] D. Junaedi, M. R. Arsyad, F. Salistia, and M. Romli, "Menguji Efektivitas Vaksinasi Covid-19 di Indonesia," *Reslaj Relig. Educ. Soc. Laa Roiba J.*, vol. 4, no. 1, pp. 120–143, 2022.
- [10] N. P. Astuti, E. G. Z. Nugroho, J. C. Lattu, I. R. Potempu, and D. A. Swandana, "Persepsi Masyarakat terhadap Penerimaan Vaksinasi Covid-19: Literature Review," *J. Keperawatan*, vol. 13, no. 3, pp. 569–580, 2021.
- [11] A. Puspasari and A. Achadi, "Pendekatan Health Belief Model Untuk Menganalisis Penerimaan Vaksinasi Covid-19 Di Indonesia," *Syntax Lit. J. Ilm. Indones.*, vol. 6, no. 8, pp. 3709–3721, 2021.
- [12] C. A. Sugianto, A. H. Rahayu, and A. Gusman, "Algoritma k-means untuk pengelompokan penyakit pasien pada puskesmas cigugur tengah," *J. Inf. Technol.*, vol. 2, no. 2, pp. 39–44, 2020.
- [13] 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.
- [14] U. R. Gurning and Mustakim, "Penerapan Algoritma K-Means dan K-Medoid untuk Pengelompokan Data Pasien Covid-19," *Build. Informatics, Technol. Sci.*, vol. 3, no. 1, p. 48–55, 2021, doi: 10.47065/bits.v3i1.1003.
- [15] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R. H. Zer, and D. Hartama, "Analisis algoritma k-medoids clustering dalam pengelompokan penyebaran covid-19 di indonesia," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 166–173, 2020.
- [16] E. Bu'ulolo and B. Purba, "Algoritma Clustering Untuk Membentuk Cluster Zona Penyebaran Covid-19," *Digit. Zo. J. Teknol. Inf. Dan Komun.*, vol. 12, no. 1, pp. 59–67, 2021.
- [17] H. Santoso, I. P. Hariyadi, and P. Prayitno, "Data Mining Analisa Pola Pembelian Produk Dengan Menggunakan Metode Algoritma Apriori," *Semasteknomedia Online*, vol. 4, no. 1, pp. 3–7, 2016.
- [18] Linada Yuliana, "Penerapan Algoritma K-Modes Clustering Untuk Pengelompokan Desa Rawan," p. 18, 2019.
- [19] D. T. Larose, *Data mining and predictive analytics*. John Wiley & Sons, 2015.
- [20] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. Dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019.
- [21] A. Ramadhan, Z. Efendi, and Mustakim, "Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.* 9, pp. 219–226, 2017.
- [22] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi, "K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir," *IEEE Access*, vol. 8, pp. 151847–151855, 2020, doi: 10.1109/ACCESS.2020.3014021.
- [23] M. Faid, M. Jasri, and T. Rahmawati, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," *Teknika*, vol. 8, no. 1, pp. 11–16, 2019, doi: 10.34148/teknika.v8i1.95.