

CLUSTERING JUMLAH POSYANDU DENGAN MENGGUNAKAN ALGORITMA K-MEANS DI KABUPATEN INDRAMAYU

Hidayatul Muhtadiin¹, Edi Tohidi²

¹Program Studi Manajemen Informatika, STMIK IKMI Cirebon , Kota Cirebon Indonesia

²Program Studi Komputerisasi Akuntansi, STMIK IKMI Cirebon , Kota Cirebon Indonesia

Email: sebastianhiday87@gmail.com¹, ²editohidi@gmail.com

Email Penulis Korespondensi : sebastianhiday87@gmail.com

Abstract—Indramayu Regency is one of the areas with major public health challenges. Efforts to develop or improve posyandu services in various locations are still not optimal because the case profile map of the areas with the highest and lowest number of posyandu is not yet known. The analysis carried out was carried out using clustering data mining techniques using the K-means algorithm. Clustering techniques group data according to the same characteristics. By using the Davies-Bouldin index value to determine the number of clusters. From the results of this research, it was found that the best clusters from clustering the number of posyandu were 6 clusters, namely cluster 0 with 28 sub-districts and 86 villages with the best number of high-level posyandu, cluster 1 with 2 sub-districts and 2 villages with a high-level number of posyandu, cluster 2 with 12 sub-districts and 20 villages with a number of low-level posyandu, cluster 3 with 26 sub-districts and 107 villages with a number of low-level posyandu, cluster 4 with 15 sub-districts and 28 villages with a number of medium-level posyandu, cluster 5 with 20 sub-districts and 63 villages with a number of medium-level posyandu .

Keyword : Data mining, Clustering , K-Means, DBI

Abstrak— Kabupaten Indramayu merupakan salah satu daerah dengan tantangan kesehatan masyarakat yang besar. Upaya pembinaan atau peningkatan layanan posyandu di berbagai lokasi masih belum optimal karena belum diketahui peta profil kasus wilyayah dengan jumlah posyandu tertinggi dan terendah. Analisis yang dilakukan dengan teknik clustering data mining dengan menggunakan algoritma K-means. Teknik clustering mengelompokkan data menurut karakteristik yang sama. Dengan menggunakan nilai indeks Davis-Bouldin untuk menentukan jumlah cluster. Dari hasil penelitian ini ditemukan bahwa cluster terbaik dari clustering jumlah posyandu terdapat 6 cluster yaitu cluster 0 dengan 28 kecamatan dan 86 desa dengan jumlah posyandu terbaik tingkat tinggi, cluster 1 dengan 2 kecamatan dan 2 desa dengan jumlah posyandu tingkat tinggi, cluster 2 dengan 12 kecamatan dan 20 desa dengan jumlah posyandu tingkat rendah, cluster 3 dengan 26 kecamatan dan 107 desa dengan jumlah posyandu tingkat rendah, cluster 4 dengan 15 kecamatan dan 28 desa dengan jumlah posyandu tingkat menengah, cluster 5 dengan 20 kecamatan dan 63 desa dengan jumlah posyandu tingkat menengah.

Kata Kunci : Data mining, Clustering , K-Means, DBI.

INTRODUCTION

Masalah kesehatan ibu dan anak selalu menjadi perhatian pemerintah. Salah satu bentuk program pemerintah untuk mengendalikan laju pertumbuhan anak usia dini adalah Pos Pelayanan Umum (Posyandu).[1] Program ini telah berperan penting dalam memberikan layanan kesehatan dasar, termasuk pencegahan penyakit dan pemantauan pertumbuhan bagi masyarakat, terutama di wilayah pedesaan dan daerah terpencil.

Setiap tahunnya, kementerian kesehatan mengumpulkan data kesehatan penduduk melalui layanan kesehatan yang tersebar di setiap daerah dan mengolah data tersebut untuk menghasilkan pemeringkatan Provinsi dan kabupaten sehat. Hasil olahan dan pemeringkatan ini menjadi acuan penting bagi pemerintah daerah.[2] Namun ada juga di bagian pelosok desa yang kurang memiliki

Posyandu, Sehingga membuat masyarakat kurang merasakan pelayanan kesehatan tersebut.

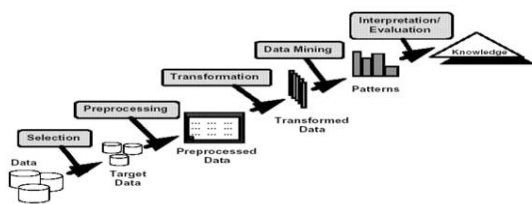
Data mining adalah proses pengumpulan dan pemrosesan data yang bertujuan untuk mengekstraksi informasi penting dari data. Salah satu metode tersebut adalah K-Means. K-Means adalah teknik pengelompokan data non-hierarki yang berupaya mempartisi data yang ada menjadi satu atau lebih cluster/grup. Ini mengelompokkan data dengan karakteristik yang sama ke dalam cluster yang sama, dan data dengan karakteristik berbeda ke dalam kelompok yang berbeda.[3] Data mining digunakan untuk menyelesaikan permasalahan jumlah Posyandu di Kabupaten Indramayu

Tujuan dari penelitian ini adalah untuk mencari cluster terbaik dari kasus jumlah Posyandu yang ada di beberapa Desa di tiap Kecamatan pada

Kabupaten Indramayu dimana dapat divisualisasikan menjadi peta profil jumlah Posyandu. Sehingga dapat memudahkan pemerintah setempat untuk memahami dan menyelesaikan masalah jumlah Posyandu yang nantinya bisa dilakukan peningkatan Layanan Kesehatan Desa di tiap Kecamatan Kabupaten Indramayu.

MATERIALS AND METHODS

Knowledge Discovery in Database (KDD) adalah kegiatan teknik pengumpulan data yang menggunakan data historis untuk menentukan kesesuaian dan merancang pola data berskala besar. Data mining menghasilkan data yang dapat digunakan untuk menghitung pilihan data tahun yang akan datang.[4] Knowledge Discovery in Databases (KDD) adalah penerapan metode ilmiah pada data mining. Dalam konteks ini, data mining adalah sebuah langkah dalam proses KDD.[5] Terdapat tahapan proses KDD sebagai berikut.



Gambar 1. Tahapan Proses KDD

Berikut adalah penjelasan tahapan proses *Knowledge Discovery in Database* :

Data Selection Dalam fase ini, memutuskan grup data dan membuat grup data target atau fokus pada sub kumpulan variabel (data sampel) yang perlu dilakukan penemuannya. Data yang diambil diurutkan dan disimpan dalam file terpisah dari database

Preprocessing Pada tahap *preprocessing* menghapus data duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan data.

Transformation Pada tahap ini, formulir data yang belum memiliki entitas unik diubah menjadi formulir data valid atau formulir data yang siap untuk proses data mining.

Data Mining Tahap *data mining* merupakan tahap yang dilakukan untuk menemukan model atau informasi unik tentang data yang diperoleh dengan menggunakan metode atau teknik algoritmik yang sesuai dengan tujuan keseluruhan proses KDD

Interpretasi/Evaluation Tahap terakhir ini bertujuan untuk mengetahui model apa saja yang diperoleh. Mengevaluasi

(mempertimbangkan) apakah pola yang diperoleh sesuai dengan hipotesis penelitian sebelumnya. Pengetahuan yang dihasilkan dari model tersebut ditampilkan dalam bentuk visualisasi

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari situs *website* Open Data Jabar tentang jumlag Posyandu. Atribut yg digunakan pada data jumlah Posyandu sebanyak 5 atribut dan 927 jumlah data yang diambil dari data jumlah Posyandu dari tahun 2019 sampai dengan tahun 2021. Dataset diperoleh dari penyedia layanan informasi data Provinsi Jawa Barat dengan *link url* <https://opendata.jabarprov.go.id/>

Teknik Pengumpulan data yang digunakan dalam penelitian ini adalah Web Scrapping yaitu data yang diperoleh melalui internet tentang jumlah poyandu yang ada di kabupaten Indramayu. Data yang diperoleh sudah dalam bentuk excel.

Dataset pada penelitian ini berisikan 4 atribur dengan 927 data yang meliputi Nama Kabupaten, Nama Kecamatan, Nama Desa, dan Jumlah Posyandu.

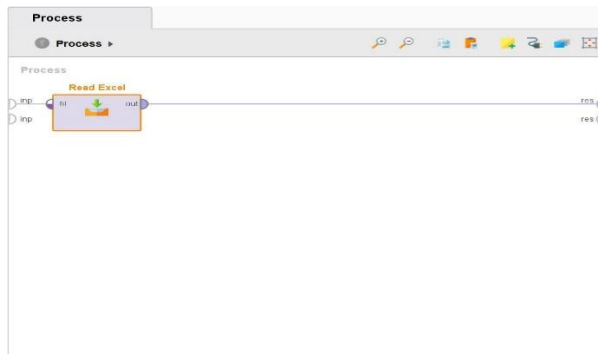
Tabel 1. Data

No	Nama Kabupaten	Nama Kecamatan	Nama Desa	Jumlah Posyandu
1	KABUPATEN INDRAMAYU	HAURGEULIS	HAURKOLOT	6
2	KABUPATEN INDRAMAYU	HAURGEULIS	HAURGEULIS	10
3	KABUPATEN INDRAMAYU	HAURGEULIS	SUKAJATI	12
4	KABUPATEN INDRAMAYU	HAURGEULIS	WANAKAYA	11
...	...	...	...	...
927	KABUPATEN INDRAMAYU	PATROL	SUKAHAJI	12

RESULTS AND DISCUSSION

A. Data Selection

Proses pertama yang dilakukan adalah seleksi data. Pada tahap ini hal yang pertama dilakukan adalah mengimport data menggunakan operator read excel. Operator read excel berguna untuk membaca data jumlah Posyandu yang ada di Kabupaten Indramayu dari tahun 2019 hingga tahun 2021.



Gambar 2. Operator Read Excel

Apabila data sudah berhasil diimportkan kedalam rapid miner menggunakan operator read excel, kemudian akan muncul preview data. Preview data menampilkan data yang akan digunakan.

Import Data - Format your columns.

☐ Replace errors with missing values

No	Nama Kabu...	Nama Keca...	Nama Desa	Jumlah Pos...	Satuan
integer	polynomial	polynomial	polynomial	integer	polynomial
1	KABUPATEN IN...	HAURGEULIS	HAURKOLOT	6	UNIT
2	KABUPATEN IN...	HAURGEULIS	HAURGEULIS	10	UNIT
3	KABUPATEN IN...	HAURGEULIS	SUKAJATI	12	UNIT
4	KABUPATEN IN...	HAURGEULIS	WANAKAYA	11	UNIT
5	KABUPATEN IN...	HAURGEULIS	KARANGTUMAR...	9	UNIT
6	KABUPATEN IN...	HAURGEULIS	KERTANEGARA	13	UNIT
7	KABUPATEN IN...	HAURGEULIS	CIPANCUNG	11	UNIT
8	KABUPATEN IN...	HAURGEULIS	MEKARUATI	11	UNIT
9	KABUPATEN IN...	HAURGEULIS	SIDODADI	9	UNIT
10	KABUPATEN IN...	HAURGEULIS	SUMBERMULYA	13	UNIT
11	KABUPATEN IN...	KROYA	TANJUNKERTA	10	UNIT
12	KABUPATEN IN...	KROYA	KROYA	10	UNIT

no problems.

Gambar 3. Operator Select Attribut

B. Preprocessing

Pada tahap ini adalah bagian dimana melakukan pengecekan pada hasil statistik data, dan tidak di temukan missing value yang artinya tidak dilakukan proses cleaning karena data yg digunakan relevan dan konsisten.

ExampleSet (Read Excel)

Name	Type	Missing	Filter (7 / 7 attributes)	Search for Attributes
No	Integer	0	1	927
Nama Desa	Nominal	0	WIRAPANJUNAN (3)	BULAK (6)
Nama Kecamatan	Nominal	0	PASEKAN (18)	JATIBARANG (45)
Nama Kabupaten	Nominal	0	KABUPATE [...] AYU (927)	KABUPATE [...] AY
Jumlah Posyandu	Integer	0	1	19
Satuan	Nominal	0	UNIT (927)	UNIT (927)
Tahun	Integer	0	2019	2021

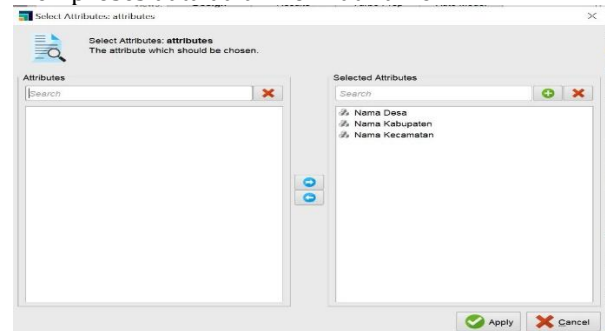
Showing attributes 1 - 7

Gambar 4. Pengecekan Missing Value

C. Transformation

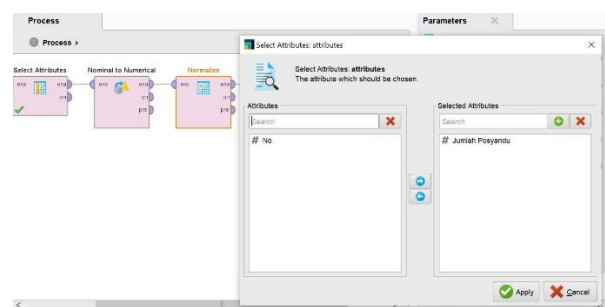
Pada tahap ini atribut-atribut yang ditentukan diubah atau disesuaikan sesuai dengan bentuk algoritma yang digunakan (dalam hal ini algoritma K-means). Dua proses dilakukan selama fase transformasi data. Yang pertama adalah menggunakan operator nominal to numerical untuk bekerja dengan atribut dalam data Penelitian, seperti nama Desa dan nama Kecamatan. Bisa

diproses karena data mining hanya dapat memproses data dalam format numerik.



Gambar 5. Transformasi Data dengan Operator Nominal to Numerical

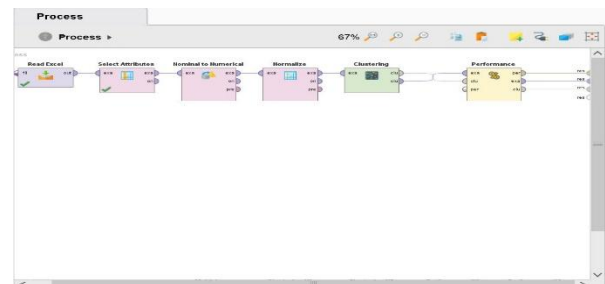
Selanjutnya pada langkah kedua yaitu melakan Normalize. Normalisasi digunakan untuk menentukan rentang nilai pada data, menyesuaikan rentang nilai sangat penting ketika berhadapan dengan atribut dari unit dan skala yang berbeda. Atribut yang digunakan dalam proses normalisasi adalah Jumlah Posyandu dan Tahun.



Gambar 6. Normalize

D. Data Mining

Setelah proses preprocessing dan Transformasi data telah di lalui, tahapan selanjutnya adalah mengcluster, dalam Penelitian ini menggunakan teknik algoritma K-Means clustering yang bisa dilihat pada gambar berikut :



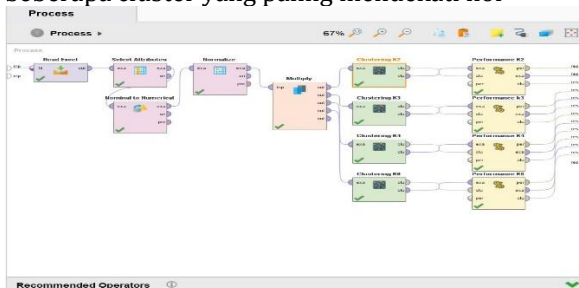
Gambar 7. Komputasi K-Means

Pada gambar 7. Menampilkan tahapan dan proses melakukan clustering menggunakan tools Rapid Miner. Dalam melakukannya dengan yang pertama

adalah menggunakan operator read excel yang digunakan untuk membaca data dalam jenis file excel, operator yang digunakan yang kedua adalah Select Attribut untuk memilih dan menentukan atribut apa saja yang akan digunakan, operator yang ketiga adalah Nominal to Numerical yang berguna untuk mengubah atribut data menjadi numerik, operator yang digunakan ke empat adalah Normalize digunakan untuk menyesuaikan rentang nilai atribut yang digunakan, perator yang digunakan ke enam adalah K-Means clustering yang digunakan untuk mengelompokan data, dan operator yang ke tujuh adalah Cluster Distance Performance yang digunakan untuk menunjukan nilai dan mengevaluasi hasil nilai Centroid Distance dan Davies Bouldin Index.

E. Interpretation/Evalution

Tahap yang terakhir adalah Interpretasi atau Evaluasi dari beberapa nilai cluster yang digunakan, dengan melihat hasil nilai Davies Bouldin Index. Dari beberapa kemungkinan yang dilakukan dengan cara mencoba menghitung nilai beberapa cluster yang paling mendekati nol



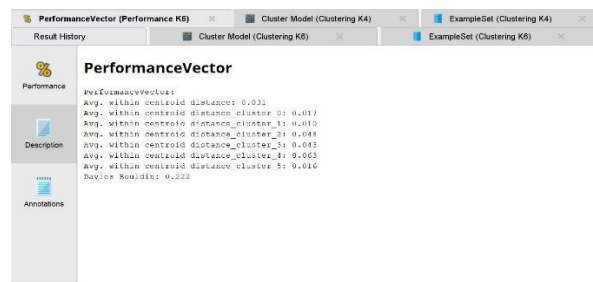
Gambar 8. Mencari Cluster Terbaik

Pada gambar 4.8 menunjukkan bagian proses penghitungan nilai menggunakan Davies Bouldin Index, yaitu dengan K2, K3, K4, dan K6 cluster yang dihitung akan diketahui nilai dari Davies Bouldin Index yang dihasilkan.

Tabel 2. Hasil Dari DBI Cluster

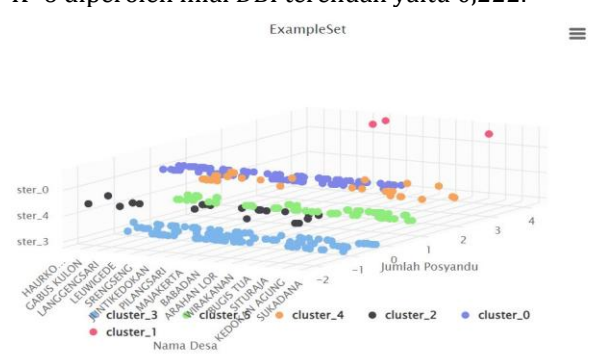
No	Cluster	Nilai Davies Bouldin Index
1	K2	0,312
2	K3	0,260
3	K4	0,270
4	K6	0,222

Berdasarkan dari Tabel 4.1 tersebut, menunjukan hasil dari terbaik adalah K=6 dengan nilai Davies Bouldin Index 0,222. Dan Performance Vector juga bisa di ketahui nilainya pada gambar berikut :



Gambar 9. Hasil Performance Vector

Dari gambar 9. Uji performance vector tersebut dapat diperoleh hasil dari 6 kelompok cluster dengan rata-rata centroid 0.031. Rata-rata centroid cluster 0 sebesar 0.017, rata-rata centroid cluster 1 sebesar 0.010, rata rata centroid cluster 2 sebesar 0.048, rata rata centroid cluster 3 sebesar 0.043, rata rata centroid cluster 4 sebesar 0.63, dan rata rata centroid cluster 5 sebesar 0.016. K=6 dipilih karena kriterianya mendekati 0 berdasarkan perhitungan yang diperoleh dari nilai DBI. Untuk K=6 diperoleh nilai DBI terendah yaitu 0,222.



Gambar 10. Visualisai Plot Scatter

Pada gambar 10. Menunjukan bahwa hasil pengelompokan jumlah posyandu dapat divisualisasikan dari plot scatter. Dimana scatter berwarna ungu menunjukan cluster 0, scatter berwarna merah menunjukan cluster 1, scatter berwarna hitam menunjukan cluster 2, scatter berwarna biru menunjukan cluster 3, scatter berwarna oren menunjukan cluster 4, dan scatter berwarna hijau menunjukan cluster 5. Dengan nilai cluster pada sumbu x dan y. Sehingga dapat disimpulkan bahwa kelompok cluster dengan cluster 0 dapat digolongkan pada jumlah Posyandu tingkat tinggi, cluster 1 dapat digolongkan pada jumlah Posyandu tingkat tinggi, cluster 2 dapat digolongkan pada jumlah Posyandu tingkat rendah, cluster 3 dapat digolongkan pada jumlah Posyandu tingkat rendah, cluster 4 dapat digolongkan pada jumlah Posyandu tingkat menengah, dan cluster 5 dapat digolongkan pada jumlah Posyandu tingkat menengah. Hasil dari pengelompokan tiap cluster adalah sebagai berikut :

Tabel 3. Cluster 0

No	Nama Kecamatan	Nama Desa	Cluster
1	GABUSWETAN	BABAKAN JAYA	cluster_0
2	GABUSWETAN	KEDOKAN GABUS	cluster_0
3	CIKEDUNG	LOYANG	cluster_0
...	...	...	...
90	PATROL	BUGEL	cluster_0

Tabel 4. Cluster 1

No	Nama Kecamatan	Nama Desa	Cluster
1	JUNTINYUAT	DADAP	cluster_1
2	GANTAR	MEKARJAYA	cluster_1

Tabel 5. Cluster 2

No	Nama Kecamatan	Nama Desa	Cluster
1	GABUSWETAN	KEDUNG DAWA	cluster_2
2	GABUSWETAN	DRUNTEN KULON	cluster_2
3	KERTASEMAYA	TEGAL WIRANGRONG	cluster_2
...	...	...	...
18	TUKDANA	KARANGKERTA	cluster_2

Tabel 6. Cluster 3

No	Nama Kecamatan	Nama Desa	Cluster
1	HAURGEULIS	HAURKOLOT	cluster_3
2	GABUSWETAN	SEKAR MULYA	cluster_3
3	GABUSWETAN	GABUS WETAN	cluster_3
...	...	...	...
108	PATROL	ARJASARI	cluster_3

Tabel 7. Cluster 4

No	Nama Kecamatan	Nama Desa	Cluster
1	HAURGEULIS	SUKAJATI	cluster_4
2	HAURGEULIS	WANAKAYA	cluster_4
3	HAURGEULIS	KERTANEGARA	cluster_4
...	...	...	...
28	PATROL	SUKAHAJI	cluster_4

Tabel 8. Cluster 5

No	Nama Kecamatan	Nama Desa	Cluster
1	HAURGEULIS	HAURGEULIS	cluster_5
2	HAURGEULIS	KARANGTUMARIS	cluster_5
3	HAURGEULIS	SIDODADI	cluster_5
...	...	...	...
63	TUKDANA	SUKADANA	cluster_5

Dari tabel diatas tersebut diperoleh bahwa Cluster terbaik dari clustering jumlah Posyandu terdapat 5 cluster yaitu cluster 0 terdapat 28 Kecamatan dan 89 Desa, cluster 1 terdapat 2 Kecamatan dan 2 Desa, cluster 2 terdapat 12 Kecamatan dan 20 Desa, cluster 3 terdapat 26 Kecamatan dan 107 Desa, cluster 4 terdapat 15 Kecamatan dan 28 Desa, cluster 5 terdapat 20 Kecamatan dan 63 Desa. Dengan nilai Avg .within centroid distance 0.031. Cluster 0 dan 1 merupakan Kecamatan dan Desa dengan jumlah Posyandu tertinggi, cluster 2 dan 3 merupakan Kecamatan dan Desa dengan jumlah Posyandu terendah, cluster 4 dan 5 merupakan Kecamatan dan Desa dengan jumlah Posyandu menengah.

CONCLUSION

Kesimpulan dari Penelitian ini, menunjukan bahwa terdapat tingkat jumlah Posyandu yang rendah, menengah dan tinggi berdasarkan data yang diolah menggunakan Rapid Miner dengan metode K-Means. Hasil perhitungan Davies Bouldin Index menunjukan bahwa penentuan cluster menggunakan 6 kelompok memiliki nilai mendekati 0.

Cluster 0 memiliki jumlah Posyandu tertinggi dengan anggota cluster sebanyak 28 Kecamatan dan 89 Desa, cluster 1 memiliki jumlah Posyandu tertinggi dengan anggota cluster sebanyak 2 Kecamatan dan 2 Desa, cluster 2 memiliki jumlah Posyandu terendah dengan anggota cluster sebanyak 12 Kecamatan dan 20 Desa, cluster 3 memiliki jumlah Posyandu terendah dengan anggota cluster sebanyak 26 Kecamatan dan 107 Desa, cluster 4 memiliki jumlah Posyandu menengah dengan anggota cluster sebanyak 15 Kecamatan dan 28 Desa, dan cluster 5 memiliki jumlah Posyandu menengah dengan anggota cluster sebanyak 20 Kecamatan dan 63 Desa.

REFERENCE

- [1] A. Aprilia, W. M. Rahmawati, and M. Hakimah, "Penentuan Kategori Status Gizi Balita Menggunakan Penggabungan Metode Klasterisasi Agglomerative Dan K-Means," *Semin. Nas. Sains dan Teknol. Terap. VII - Inst. Teknol. Adhi Tama Surabaya*, pp. 595–600, 2019.
- [2] N. Atthina and L. Iswari, "Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat Kesehatan Daerah dengan Metode K-Means," *Semin. Nas. Apl. Teknol. Inf. Yogyakarta*, pp. 21–2014, 2022.
- [3] D. Triyansyah and D. Fitriana, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *J. Telekomun. dan Komput.*, vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [4] M. Syahril, K. Erwansyah, and M. Yetri, "Penerapan Data Mining Untuk Menentukan Pola Penjualan Peralatan Sekolah Pada Brand Wigglo Dengan Menggunakan Algoritma Apriori," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 3, no. 1, p. 118, 2020, doi: 10.53513/jsk.v3i1.202.
- [5] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.